Electronic Communication Systems

12

Semiconductor Microwave Devices and Circuits

No segment of the microwave field has had more research devoted to it, over the past three decades, than the field of solid-state devices and circuits. This has resulted in a tremendous proliferation of, and improvements in, semiconductor devices for microwave amplification, oscillation, switching, limiting, frequency multiplication and other functions. For the systems designer, the result of these continuing improvements has been greater flexibility, improved performance, generally greater reliability, reduced sizes and power requirements, and importantly the ability to produce some systems that would not otherwise have been possible.

It would be entirely feasible to write a large book on each of the major sections of this chapter. In this chapter we will explain the basic principles of each type of device, to discuss its practical aspects and applications, to describe and show its appearance, and to indicate its state-of-the-art performance figures. Different devices that may be used for similar purposes will be compared from a practical point of view. A number of explanations will be deliberately simplified because of the complex nature of the material.

The chapter begins with an explanation of certain passive microwave circuits, notably *microstrip, stripline* and *surface acoustic wave (SAW)* components. They are not semiconductor devices themselves, but since they are often used in conjunction with solid-state microwave devices, this is a convenient place to review them.

We then continue with a presentation of microwave transistors, both bipolar and field-effect. As with microwave triodes in the preceding chapter, it will be assumed that students already understand how transistors work. We will then discuss their high-frequency limitations and what makes microwave transistors different in construction and behavior from lower-frequency ones. The section concludes with an introduction to microwave integrated circuits.

The next section is devoted to varactor diodes. These are diodes whose capacitance is linearly variable with the change in applied bias. This property makes the diodes ideal for electronic tuning of oscillators and for low-loss frequency multiplication. Another important application of varactors is in *parametric amplifiers,* which form the next major portion of the chapter. Extremely low-noise amplification of (microwave) signals can be obtained by a suitable variation of a reactive parameter of an RLC circuit. Varactor diodes fit the bill, since their capacitance parameter is easily variable.

Tunnel diodes and their applications are the next topic studied. They are diodes which, under certain circumstances, exhibit a negative resistance. It will be shown that this results in their use as amplifiers and oscillators. Tunnel diodes will be used as an exam-

ple of how amplification is possible with a device that has negative resistance.

The *Gunn effect* and *Gunn diodes,* so-called after their inventor, are discussed next. These are devices in which negative resistance is obtained as a *bulk* property of the material used, rather than a junction property. Gunn diodes are now very common medium-power oscillators for microwave frequencies, with a host of applications that will be covered.

Another class of power devices depends on *controlled avalanche* to produce microwave oscillations or amplification. The *IMPATT* and *TRAPATT* diodes are the most commonly used, and both are discussed in the next section of the chapter. They are followed in the next-to-last section by an explanation of the *Schottky barrier* and *PIN diodes,* used for mixing/detection and limiting/switching, respectively.

The final topic covered is the amplification of microwaves or light by means of the quantum-mechanical effect of stimulated emission of radiation. The topic covers masers, lasers and a number of other optoelectronic devices.

OBJECTIVES

Upon completing the material in Chapter 12, the student will be able to:

Understand the theory and application of stripline and microstrip circuits and SAW devices.

Explain the construction, limitation, and performance characteristics of microwave integrated circuits, transistors, and diodes.

Define the term *maser.*

Discuss the differences between masers and lasers.

12-1 PASSIVE MICROWAVE CIRCUITS

Transmission lines and waveguides were invented at the time of, and used in conjunction with, microwave electron tubes such as those discussed in the preceding chapter. They are still so used at medium and high powers. Again, being low-loss, they are used at low powers where significant distances are traversed, as in connecting antennas to receivers. However, transmission lines are considerably bulkier than semiconductor microwave devices, and consequently their use would prevent the reduction in circuit size and weight which would otherwise be obtainable. *Stripline* and *microstrip* have been developed and are used for circuit interconnections with solid-state devices. They may also be used for passive components, as can another class of devices, using SAW principles. *Microwave integrated circuits* (MICs) are not uncommon and have many applications.

12-1.1 Stripline and Microstrip Circuits

Stripline and *microstrip* are physically related to transmission lines but are covered here because they are microwave circuits used in conjunction with semiconductor

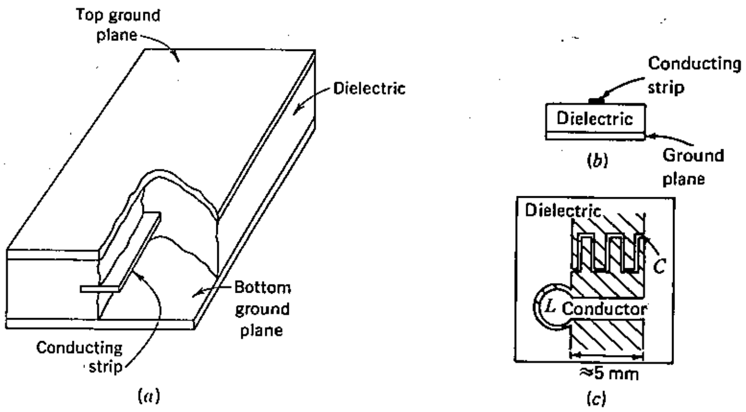

FIGURE 12-1 **(*a*) Stripline; (*b*) microstrip cross section; (*c*) microstrip LC circuit.**

microwave devices. As illustrated in Figure 12-1, *stripline* consists of flat *metallic ground planes*, separated by a thickness of dielectric in the middle of which a thin metallic strip has been buried. The conducting strip in *microstrip* is on top of a layer of dielectric resting on a single ground plane. Typical dielectric thicknesses vary from 0.1 to 1.5 mm, although the metallic strip may be as thin as 10 μm.

Stripline and microstrip were developed as an alternative conducting medium to waveguides and are now used very frequently in a host of microwave applications in which miniaturization has been found advantageous. Such applications include receiver front ends, low-power stages of transmitters and low-power microwave circuitry in general.

Stripline is evolved from the coaxial transmission line. It may be thought of as flattened-out coaxial line in which the edges have been cut away. Propagation is similarly by means of the TEM (transverse electromagnetic) mode as a reasonable approximation. Microstrip is analogous to a parallel-wire line, consisting of the top strip and its image below the ground plane. The dielectric is often Teflon, alumina or silicon. It is possible to use several independent strips with the same ground planes and dielectric, for both types of circuits. Semiconductor microwave devices are often packaged for direct connection to stripline or microstrip.

As was shown in Chapter 10, waveguides are used not only for interconnection but also as circuit components. The same applies to stripline and microstrip (and indeed to coaxial lines). Figure 12-1*c* shows a microstrip *LC* circuit—typical capacitances possible are up to 1 pF, and typical inductances up to 5 nH. The stripline version would be very similar, with just a covering of dielectric and a second ground plane. Transformers can be made similar to the single-turn coil shown, and passive filters and couplers may also be fabricated. Resistances are obtained by using a patch of high-resistance metal such as Nichrome, instead of the copper conductor. Ferrite may be readily blended into such circuits, and so isolators, circulators and duplexers (all described in Chapter 10) are quite feasible. Figure 10-43 shows the construction of a ferrite stripline circulator.

Microstrip has the advantage over stripline in being of simpler construction and easier integration with semiconductor devices, lending itself well to printed-circuit and thin-film techniques. On the other hand, there is a far greater tendency with microstrip to radiate from irregularities and sharp corners. Thus there is a lower isolation between adjoining circuits in microstrip than in stripline. Finally, both Q and power-handling ability are lower with microstrip.

In comparison with waveguides (and coaxial lines), stripline has two significant advantages; reduced bulk and greater bandwidth. The first of these goes without saying, while the second is due to a restriction in waveguides. In practice, these are used over the 1.5:1 frequency range, limited by cutoff wavelength at the lower end and the frequency at which higher modes may propagate at the upper end. There is no such restriction with stripline, and so bandwidths greater than 2:1 are entirely practicable. A further advantage of stripline, as compared with waveguides, is greater compatibility for integration with microwave devices, especially semiconductor ones. On the debit side, stripline has greater losses, lower Q and much lower power-handling capacity than waveguides. Circuit isolation, although quite good, is not in the waveguide class. The final disadvantage of stripline (and consequently of microstrip) is that components made of it are not readily adjustable, unlike their waveguide counterparts.

Above about 100 GHz, stripline and microstrip costs and losses rise significantly. However, at frequencies lower than that, these circuits are very widely used, particularly at low and medium powers.

12-1.2 SAW Devices

Surface acoustic waves may be propagated on the surfaces of solid piezoelectric materials, at frequencies in the VHF and UHF regions. Devices employing SAW principles were first discussed in the late 1960s, then moved out of the laboratory in about 1974, and since about 1978 have found many applications as passive components in the low microwave range.

The application of an ac voltage to a plate of quartz crystal will cause it to vibrate and, if the frequency of the applied voltage is equal to a mechanical resonance frequency of the crystal, the vibrations will be intense. Because quartz is piezoelectric, all mechanical vibrations will be accompanied by electric oscillations at the same frequency. The mechanical vibrations can be made very stable in frequency, and consequently piezoelectric crystals find many applications in stable oscillators and filters. As the desired frequency of operation is raised, so quartz plates must be made thinner and thus more fragile, so that crystal oscillators are not normally likely to operate at fundamental frequencies much in excess of 50 MHz. It is possible to multiply the output frequency of an oscillator almost indefinitely (see also Section 12-3.3), but inconvenience would be avoided if multiplication were unnecessary. This may be done with SAW resonators, which employ thin lines etched on a metallic surface electrodeposited on a piezoelectric substrate. The etching is performed by using photolithography or electron beam techniques, while the most commonly used piezoelectric materials are quartz and lithium niobate.

A simplified sketch of a typical interdigitated SAW resonator is shown in Figure 12-2. Traveling waves in both directions result from the application of an RF voltage

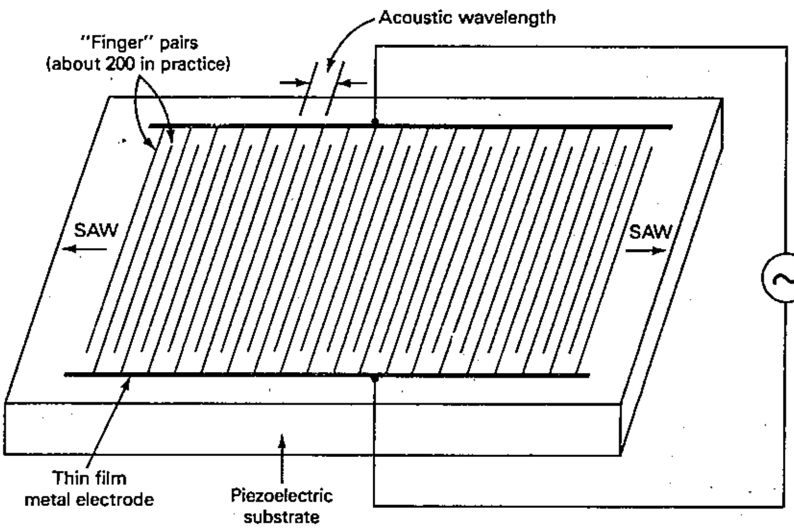

FIGURE 12-2 Basic surface acoustic wave (SAW) resonator.

between the two electrodes, but the resulting standing wave is maintained adequately only at the frequency at which the distance between adjoining "fingers" is equal to an (acoustic) wavelength, or a multiple of a wavelength along the surface of the material. As with other piezoelectric processes, an electric oscillation accompanies the mechanical surface oscillation.

If the device is used as a filter, only those frequencies that are close to the resonant frequency of the SAW resonator will be passed. Because the mechanical Q is high (though not quite as high as that of a quartz crystal being used as a standard resonator), the SAW device is a narrowband bandpass filter. To use the SAW resonator to produce oscillations, one need merely place it, in series with a phase-shift network, between the input and output of an amplifier. The phase shift is then adjusted so as to provide positive feedback, and the amplifier will produce oscillations as the frequency permitted by the SAW resonator.

There is no obvious lower limit to the operating frequency of a SAW resonator, except that it is unlikely to be used below about 50 MHz, because at such frequencies straightforward crystal oscillators can be used. The upper frequency limit is governed by photoetching accuracy. Because wavelength = v/f and the velocity of the acoustic wave is approximately 3000 m/s, it is easy to calculate that the finger separation at 5 GHz should be 0.6 μm, and the fingers themselves must be thinner still. In consequence, 5 GHz represents the current upper limit of SAW resonator operation.

12-2 TRANSISTORS AND INTEGRATED CIRCUITS

All devices that try to overcome the limitations of transit time eventually lose. This was stated in connection with vacuum tubes near the beginning of Chapter 11 and applies equally here. With tubes and transistors, useful operation can be extended into the microwave region.

12-2.1 High-Frequency Limitations

As stated, transistors suffer from high-frequency limitations. These are of a twofold nature. On the one hand, there are the same difficulties as those encountered with tubes (Section 11-1.1). On the other hand, there is some difficulty in specifying accurately the performance of microwave transistors in a manner which would make it relatively easy for the equipment designer to use them.

Limitations If one has become familiar with the limitations of vacuum tubes at high frequencies, it is relatively easy to predict the limitations of transistors. Thus, capacitances between electrodes play an important part in determining high-frequency response. Both current gains, α and β, eventually acquire reactive components which make both complex at first and eventually unusable. Interelectrode capacitances in bipolar transistors depend also on the width of the depletion layers at the junctions, which in turn depend on bias. The situation is somewhat more complex than with tubes, whose interelectrode capacitances are not so bias-dependent. The difficulty here is not that the transistor has a poorer high-frequency response; quite the opposite. It is simply a greater difficulty in finding parameters with which to describe the behavior so as to give a meaningful picture to the circuit designer.

Electrode inductances have more or less the same nuisance value as with tubes, but since transistors are smaller, electrode leads are shorter. Thus suitable geometry and the use of low-inductance packages go a long way toward reducing the effects of lead inductance.

The effect of transit time is identical to that in tubes, although its actual operation is somewhat different. The smaller distances traveled in transistors are counterbalanced by the slower velocities of current carriers, but overall the maximum attainable frequencies are somewhat higher than for tubes. In traveling across a bipolar transistor, the holes or electrons drift across with velocities determined by the ion mobility [basically higher for germanium (Ge) and gallium arsenide (GaAs) than silicon (Si)] the bias voltages and the transistor construction. We first find majority carriers suffering an emitter delay time, and then the injected carriers encounter the base transit time, which is governed by the base thickness and impurity distribution. The collector depletion-layer transit time comes next. This is governed mainly by the limiting drift velocity of the carriers (if a higher voltage were applied, damage might result) and the width of the depletion layer (which is heavily dependent on the collector voltage). Finally, electrons or holes take some finite time to cross the collector, as they did with the emitter.

Specification of performance Several methods are used to describe and specify the overall high-frequency behavior of RF transistors. Older specifications showed the alpha and beta cutoff frequencies, respectively $f_{\alpha b}$ and $f_{\alpha e}$. The first is the frequency at which α, the common-base current gain, falls by 3 dB, and the second applies similarly to β, the common-emitter current gain. The two figures are simply interconnected. Since we know that

$$\beta = \frac{\alpha}{1-\alpha} \tag{12-1}$$

it follows that, for the usual values of β,

$$f_{\alpha e} = \frac{f_{\alpha b}}{\beta} \tag{12-2}$$

These frequencies are no longer commonly in use. They have been replaced by f_T, the (current) gain-bandwidth frequency. This may simply be used as a gain-bandwidth product at low frequencies or, alternatively, as the frequency at which β falls to unity, i.e., the highest frequency at which *current* gain may be obtained. It is very nearly equal to $f_{\alpha b}$ in most cases, although it is differently defined.

Up to a point, f_T is proportional to both collector voltage and collector current and reaches its maximum for typical bipolar RF transistors at V_{ce} = 15 to 30 V and I_c in excess of about 20 mA. This situation is brought about by the higher drift velocities and therefore shorter transit times corresponding to the higher collector voltage and current.

Finally, there is one last frequency of interest to the user of microwave transistors. This is the maximum possible frequency of oscillation, f_{max}. It is higher than F_T because, although β has fallen to unity at this frequency, power gain has not. In other words, at $\beta = 1$ output impedance is higher than input impedance, voltage gain exists, and both regeneration and oscillation are possible. Although the use of transistors above the beta cutoff frequency is certainly possible and very often used in practice, the various calculations are not as easy as at lower frequencies. The transistor behaves as both an amplifier and a low-pass filter, with a 6 dB per octave gain drop above a frequency whose precise value depends on the bias conditions.

To help with design of transistor circuits at microwave frequencies, scattering-(S) parameters have been evolved. These consider the transistor as a two-port, four-terminal network under matched conditions. The parameters themselves are the forward and reverse transmission gains, and the forward and reverse reflection coefficients. Their advantage is relatively easy measurement and plotting on the Smith chart.

12-2.2 Microwave Transistors and Integrated Circuits

Silicon bipolar transistors were first on the microwave scene, followed by GaAs field-effect transistors. Indeed, FETs now have noticeably lower noise figures, and in the C band and above they yield noticeably higher powers. A description of microwave transistor constructions and a discussion of their performance now follow.

Transistor construction The various factors that contribute to a maximum high-frequency performance of microwave transistors are complex. They include the already mentioned requirement for high voltages and currents, and two other conditions. The first of these is a small electrode area to reduce interelectrode capacitance. The second is very narrow active regions to reduce transit time.

For bipolar transistors, these requirements translate themselves into the need for a very small emitter junction and a very thin base. Silicon planar transistors offer the best bipolar microwave performance. Fabrication difficulties, together with the excellent performance of GaAs FETs, have prevented the manufacture of GaAs bipolars. Epitaxial diffused structures are used, giving a combination of small emitter area and large emitter edge. The first property gives a short transit time through the emitter,

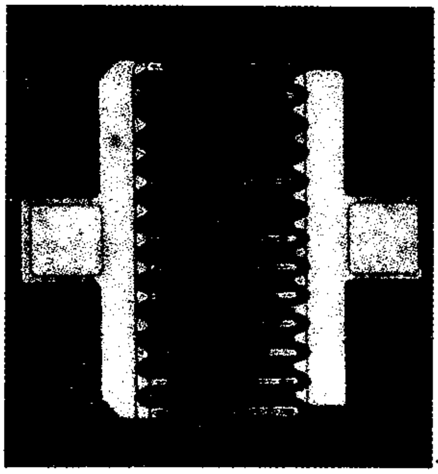

FIGURE 12-3 Geometry of an interdigitated planar microwave transistor. *(Courtesy of Texas Instruments, Inc.)*

and the second a large current capacity. The *interdigitated* transistor, shown in Figure 12-3, is by far the most common bipolar in production. The transistor shown has a base and emitter layout that is similar to two hands with interlocking fingers, hence its name. The chip illustrated has overall dimensions (less contacts) of about 70 × 70 μm; the emitter contact is on the left, the base on the right and the collector underneath. The thickness of each emitter (and base) "finger" in the transistor shown is 0.5 μm. This yields values of f_{max} in excess of 20 GHz; 0.25-μm geometries have been proposed.

The most common microwave FET uses a *Schottky-barrier gate* (i.e., a metal-semiconductor one; see also Section 12-8.2). Figure 12-4 demonstrates why this device is also known as a *MESFET*. The cross section shows it to be of mesa construction. The top metallic layer has been etched away, as has a portion of the n-type GaAs

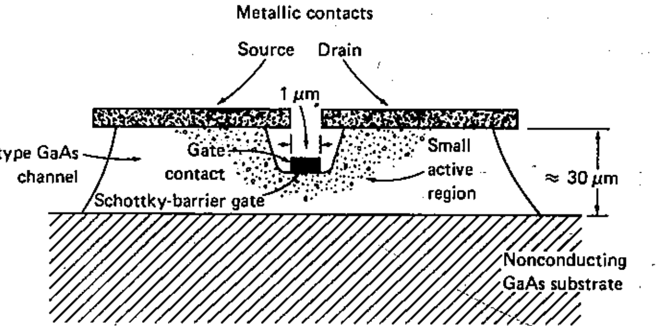

FIGURE 12-4 Construction of microwave mesa field-effect transistor (MESFET) chip, with a single Schottky-barrier gate.

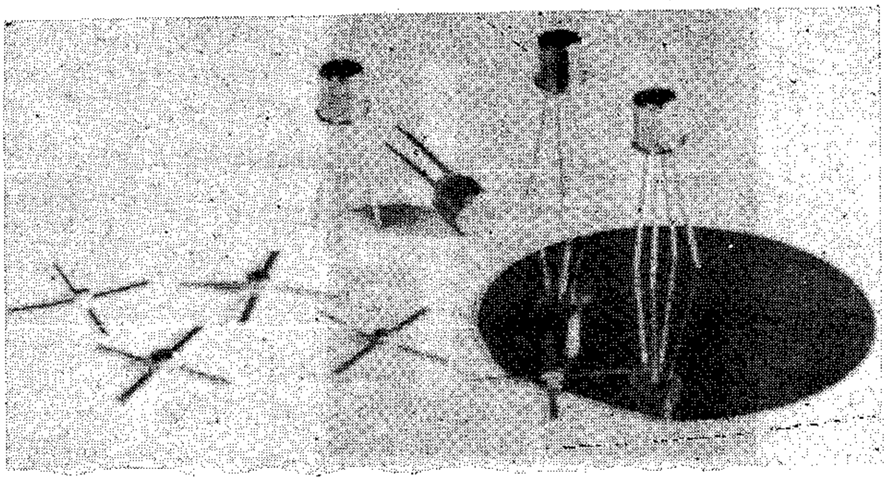

FIGURE 12-5 Microwave transistor package types. Top: T0-72 can; bottom: Stripline (beam-lead) package. *(Courtesy of Avantek, Inc.)*

semiconductor underneath. The metallic Schottky-barrier gate stripe is deposited in the resulting groove. It has a typical length of 1 μm (the normal range is 0.5–3 μm). The width of the gate is not shown in the cross section; 300–2400 μm is a typical range. Dual-gate GaAs FETs are also available, in which the second gate may be used for the application of AGC in receiver RF amplifiers. It should be mentioned that values of f_{max} in excess of 100 GHz are currently achievable.

Packaging and circuits Two typical methods of packaging microwave transistor chips are shown in Figure 12-5. The Avantek stripline package at the bottom has a body thickness of 1 mm and a diameter just under 5 mm. The TO-72 can at the top has a 7½-mm diameter and much the same height. The TO-72 package is available for frequencies up to about 2 GHz, especially for silicon bipolar transistors. The stripline packages are used for higher frequencies, up to about 30 GHz, for bipolars or FETs. For still-higher frequencies or large bandwidths, the transistor chips are bonded directly to the associated circuitry.

12-2.3 Microwave Integrated Circuits

Because of the inherent difficulties of operation at the highest frequencies, MICs took longer to develop than integrated circuits at lower frequencies. However, by the mid-1970s, *hybrid* MICs had become commercially available, at first with sapphire substrates and subsequently with (insulator) gallium arsenide substrates. In these circuits, thick or thin metallic film was deposited onto the substrate, and the passive components were etched onto the film, while the active components, such as transistors and diodes, were subsequently soldered or bonded onto each chip. In the early 1980s,

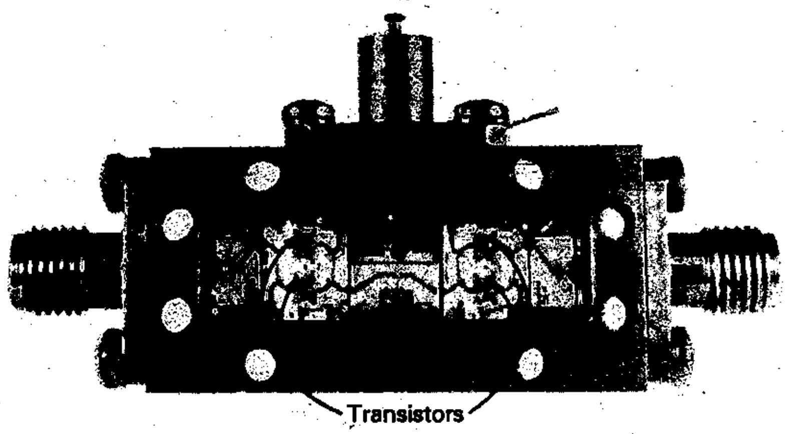

FIGURE 12-6 Hybrid GaAs FET MIC amplifier. ***Note:*** **Hermetically sealed cover removed.** *(Courtesy of Avantek, Inc.)*

however, *monolithic* MICs became commercially available. In these circuits, all the components are fabricated on each chip, using metallic films as appropriate for passive components and injection doping of the GaAs substrate to produce the requisite diodes and FETs. In view of the size reduction initially available from monolithic MICs, it appeared at first that they would completely take over the field, but significant improvements were made in hybrid circuits, with a consequent resurgence of their use. It would appear that the two types will be used side by side for the foreseeable future.

A typical hybrid MIC amplifier is illustrated in Figure 12-6. This is an Avantek miniature GaAs FET hybrid MIC, with overall dimensions (including connectors and dc power feedthrough) of about 40 × 20 × 4 mm—its volume is thus under 0.2 in^3. The two-stage amplifier produces an output of 10 mW, with a gain of 9 dB and a noise figure of 8 dB, over the very wide frequency range of 6 to 18 GHz. It is seen that the two modules on either side of center are identical balanced amplifiers, with the two transistors located above each other in the middle of each module as indicated. In a working amplifier, a lid is welded on, dry nitrogen is pumped in, and the amplifier is hermetically sealed.

A Texas Instruments monolithic MIC chip is shown in Figure 12-7. This is a high-gain four-stage GaAs FET power amplifier developed for satellite communications. Although the chip measures only I × 5.25 × 0.15 mm, it produces an output of 1.3 W at 7.5 GHz, with a good frequency response from 6.5 to 8 GHz and an efficiency of 30 percent; the gain is 32 dB. The gate widths range from 300 μm for the input FET to 2400 μm for the output FET. Silicon nitride capacitors are used, and a fair amount of gold plating is used to reduce resistance.

12-2.4 Performance and Applications of Microwave Transistors and MICs

The power and noise capabilities of microwave transistors and MICs have been improving spectacularly over more than a decade, with good improvements in bandwidth and efficiency over the same period.

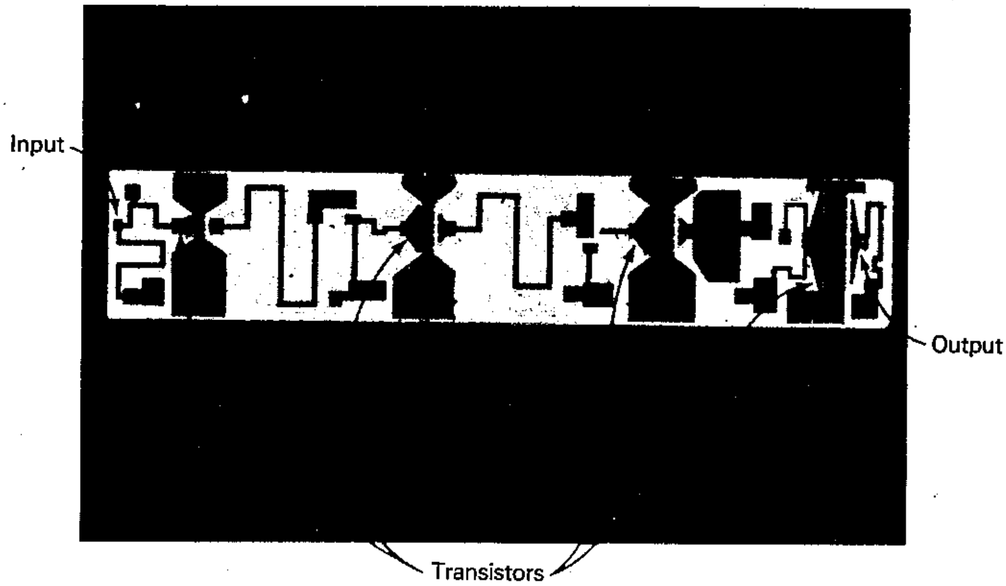

FIGURE 12-7 GaAs FET monolithic MIC four-stage high-gain power amplifier. *(Courtesy of Texas Instruments, Inc.)*

Bipolar transistors are available for frequencies up to about 8 GHz, where power devices produce up to about 150 mW output, while low-noise transistors have noise figures of the order of 14 dB. Neither is as good as the corresponding figure for GaAs FETs. However, bipolars do very well at lower microwave frequencies: transistors such as the Avantek ones shown in Figure 12-5 produce noise figures as low as 2.8 dB at 4 GHz and 1.8 dB at 2 GHz, and power bipolars can produce over 1 W per transistor at 4 GHz.

GaAs FETs are available, as discrete transistors and/or MICs, right through the Ka band (26.5 to 40 GHz) and are becoming available for higher frequencies. Powers of several watts per transistor are available up to 15 GHz, and hundreds of milliwatts to 30 GHz. Noise figures below 1 dB are attainable at 4 GHz and are still only about 2 dB at 20 GHz. The noise figures of amplifiers, be they bipolar or FET, are not as good as those of individual transistors. The major reason for this is the low gain per stage, typically 5 to 8 dB at X band (8 to 12.5 GHz). As explained in Section 2-3.2, the noise of transistors and components beyond the input stage makes a significant contribution to the total noise.

As has been mentioned, FETs have the advantage over bipolars at the highest frequencies because they are able to use GaAs, which has a higher ion mobility than silicon. They also have higher peak electron velocities, the two advantages providing a faster transit time and lower dissipation. FETs are thus able to work at higher frequencies, with higher gain, lower noise and better efficiency. Other semiconductor materials currently being investigated as potentially useful at microwave frequencies, because of possible advantages in electron mobility and drift velocity over gallium arsenide, include gallium-indium arsenide (GaInAs).

With such excellent performance, transistor amplifiers (and oscillators) have found many microwave applications, especially as their prices have fallen. The advantages of transistors over other microwave devices include long shelf and working lives, small size and electrode voltages, and low power dissipation together with good efficiencies, of the order of 40 percent. The noise figures and bandwidths are also excellent. Computer control of design and manufacture has resulted in good reliability and repeatability of characteristics for both field-effect and bipolar transistors.

Low-noise transistor amplifiers are employed in the front ends of all kinds of microwave receivers, for both radar and communications. That is, unless the requirement is for extremely low noise, in which case transistors are used to amplify the output of more exotic RF amplifiers (treated later in this chapter). The application for microwave power transistors is as power amplifiers or oscillators in a variety of situations. For example, they serve as output stages in microwave links, driver amplifiers in a wide range of high-power transmitters (including radar ones), and as output stages in broadband generators and phased array radars (see Section 16-3).

12-3 VARACTOR AND STEP-RECOVERY DIODES AND MULTIPLIERS

Step-recovery diodes are junction diodes which can store energy in their capacitance and then generate harmonics by releasing a pulse of current. They are very useful as microwave frequency multipliers, sometimes by very high factors. The *varactor*, or variable capacitance diode, is also a junction diode. It has the very useful property that its junction capacitance is easily varied electronically. This is done simply by changing the reverse bias on the diode. This single property makes this diode one of the most useful and widely employed of all microwave semiconductor devices.

12-3.1 Varactor Diodes

Varactor diodes were first used in the early 1950s as simple voltage-variable capacitances and later for frequency modulation of oscillators. They thus represent a very mature semiconductor microwave art. As materials and construction improved, so did the maximum operating frequencies, until the stage has now been reached where the most common applications are in tuning, in microwave frequency multipliers and in the very low-noise microwave parametric amplifiers (see Section 12-4).

Operation When reverse-biased, almost any semiconductor diode has a junction capacitance which varies with the applied back bias. If such a diode is manufactured so as to have suitable microwave characteristics, it is then usually called a *varactor diode;* Figure 12-8 shows its essential characteristics. Apart from the fact that the capacitance variation must be appreciable in a varactor diode, it must be capable of being varied at a microwave rate, so that high-frequency losses must be kept low. The basic way in which such losses are reduced is the reduction in the size of the active parts of the diode itself.

In a diffused-junction diode, the junction is depleted when reverse bias is applied, and the diode then behaves as a capacitance, with the junction itself acting as a dielectric between the two conducting materials. The width of the depletion layer

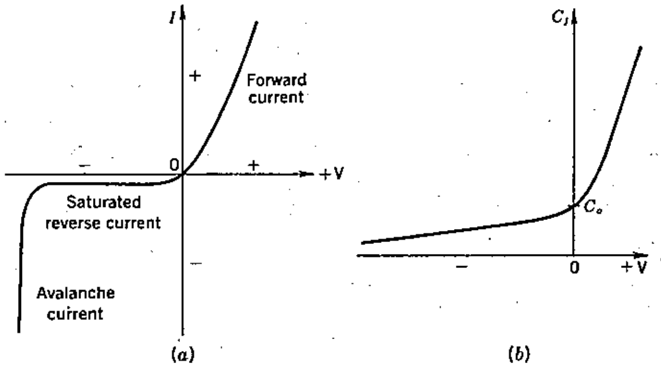

FIGURE 12-8 Varactor diode characteristics. (*a*) Current vs. voltage; (*b*) junction (depletion layer) capacitance vs. voltage.

depends on the applied bias, and the capacitance is naturally inversely proportional to the width of this layer; it may thus be varied with changes in the bias. This is shown in Figure 12-8*b*, where C_0 represents the junction capacitance for zero bias voltage. Finally, as with all other diodes, avalanche occurs with very high reverse bias. Since this is likely to be destructive, it forms a natural limit for the useful operating range of the diode.

Materials and construction Diffused-junction mesa silicon diodes were used originally at microwave frequencies, but by now they have to a large extent been supplanted by gallium arsenide varactors. Figure 12-9 shows a varactor diode made of gallium arsenide. GaAs has such advantages as a higher maximum operating frequency (up to nearly 1000 GHz) and better functioning at the lowest temperatures (of the order of −269°C, as in parametric amplifier applications). Both advantages are due mainly to the higher mobility of charge carriers exhibited by gallium arsenide.

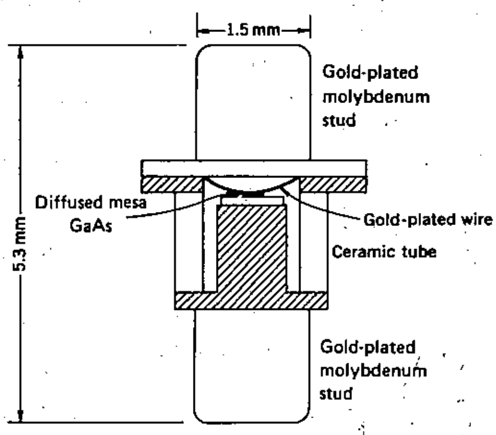

FIGURE 12-9 Varactor diode construction.

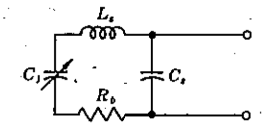

FIGURE 12-10 Varactor diode equivalent circuit.

Characteristics and requirements Above all, the varactor diode (no matter how it is made or what it is made from) is a diode, i.e., a rectifier. The diode conducts normally in the forward direction, but the reverse current saturates at a relatively low voltage (as Figure 12-8*a* shows) and then remains constant, eventually rising rapidly at the avalanche point. For varactor applications, the region of interest lies between the reverse saturation point, which gives the maximum junction capacitance, and a point just above avalanche, at which the minimum diode capacitance is obtained. Conduction and avalanche are thus seen to be the two conditions which limit the reverse voltage swing and therefore the capacitance variation.

Within the useful operating region, the varactor diode at high frequencies behaves as a capacitance in series with a resistance. At higher frequencies still, the stray lead inductance becomes noticeable, and so does the stray fixed capacitance between the cathode and anode connections. The equivalent circuit diagram of Figure 12-10 then applies. For a typical silicon varactor, $C_0 = 25$ pF, $C_{\min} = 5$ pF, $R_b = 1.3\ \Omega$, $C_s = 1.4$ pF, and $L_s = 0.013\ \mu$H.

To be suitable for parametric amplifier service, as will be seen in Section 12-4, a varactor diode should have a large capacitance variation, a small value of minimum junction capacitance and the lowest possible value of series resistance R_b (to give low noise). For harmonic generation, much the same requirements apply (although possibly the low value of R_b is a little less important), but now power-handling ability assumes a greater significance. Base resistance and minimum junction capacitance are largely tied to each other, so that these two requirements can be satisfied only in a compromise fashion. The *resistive cutoff frequency* is often used as a figure of merit; it is given by

$$f_c = \frac{1}{2\pi R_b C_{\min}} \qquad \textbf{(12-3)}$$

Values of f_c well over 1000 GHz are available from gallium arsenide varactors. However, this does not mean that varactors may be operated at such high frequencies. The f_c is measured at a relatively low frequency (e.g., 50 or 500 MHz). It is a figure of merit, a convenient way of relating base resistance and minimum junction capacitance. Operation at frequencies much above $f_c/10$ is inadvisable, because at such frequencies there is a gradual increase in base resistance, partly through the skin effect. Consequently the diode Q drops, and the result is increased noise in parametric amplifiers or increased dissipation (lowered efficiency) in frequency multipliers.

Frequency multiplication mechanism It was shown in Equation (4-3) that the output current resulting from the application of an ac voltage to a nonlinear resistance is not merely proportional to this voltage. In fact, coefficients of nonlinearity exist, and the

output current is thus in part dependent on the square, cube and higher powers of the input voltage. Equation (4-8) showed that, if the square term is taken into consideration, the output voltage contains the second harmonic of the input current. Had higher nonlinearity terms been included in the expansion, third and higher harmonics of the input would have been shown to be present in the output of such a nonlinear resistance.

Unfortunately, this type of frequency multiplication process is not very efficient, because the coefficient of nonlinearity is not usually very large. However, if Equation (4-3) is applied to a *nonlinear impedance,* the result still holds. Moreover, if this impedance is a *pure reactance,* the frequency multiplication process may be 100 percent efficient in theory.

Since the capacitance of a varactor diode varies with the applied reverse bias, the diode acts as a nonlinear capacitance (i.e., a nonlinear capacitive reactance). The varactor diode is consequently a very useful device, especially since it will operate at frequencies much higher than the highest operating frequencies of transistor oscillators.

12-3.2 Step-Recovery Diodes

A step-recovery diode, also known as a *snap-off varactor,* is a silicon or gallium arsenide *p-n* junction diode, of a construction similar to that of the varactor diode. It is an epitaxial diffused junction diode, designed to store charge when it is conducting with a forward bias. When reverse bias is applied, the diode very briefly discharges this stored energy, in the form of a sharp pulse very rich in harmonics. The duration of this pulse is typically 100 to 1000 ps, depending on the diode design. This snap time must in practice be shorter than the reciprocal of the output frequency; for example, for an output frequency of 8 GHz, snap time should be less than $T = \frac{1}{8} \times 10^{-9} = 1.25 \times 10^{-10} = 125$ ps.

As will be shown in the next section, a step-recovery diode is biased so that it conducts for a portion of the input cycle. The depletion layer of the junction is charged during this period. When the input signal changes polarity and the diode is biased off, it then produces this sharp pulse, which is very rich in harmonics. All that is then needed in the output is a tuned circuit operating at the wanted harmonic, be it the second or the twentieth. If the circuit is correctly designed, efficiencies well in excess of $1/n$ are possible, where n is the frequency multiplication factor. This means that feeding 12 W at 0.5 GHz to a snap-off varactor may result in decidedly more than 1.2 W out at 5 GHz.

It is also possible to use these diodes without a tuned output circuit, to produce multiple harmonics in so-called "comb generators." Also possible is the stacking of two or more step-recovery (or varactor) diodes in the one package, to provide a higher power-handling capacity.

12-3.3 Frequency Multipliers

Practical circuits A typical multiplier chain is shown in Figure 12-11. The first stage is a transistor crystal oscillator, operating in the VHF region, and this is the only circuit in the chain to which dc power is applied. The next stage is a step-recovery multiplier

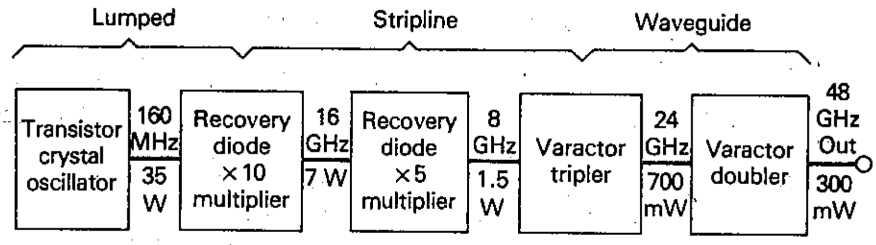

FIGURE 12-11 Step-recovery/varactor diode frequency multiplier with typical powers and frequencies shown.

by 10, bringing the output into the low-GHz range. This multiplier is likely to have lumped input circuitry and stripline or coaxial output. With 10 × multiplication, the efficiency will be of the order of 20 percent, as shown in Figure 12-11. Another snap-off 5 × multiplier now brings the output into the X band, with comparable efficiency. Normal varactors are used from this point onward. The reason is an increasing difficulty, beyond the X band, in constructing step-recovery diodes with snap-times sufficiently short to meet the $1/f_{out}$ criterion.

The circuit of Figure 12-12 shows a simple frequency tripler, which could be varactor or step-recovery. It can also be taken as the equivalent of a higher frequency stripline or cavity tripler. Note that the diode bias is provided by resistor R_B, in a leak-type arrangement. For correct operation of a snap-off varactor multiplier, the value of the resistance is normally between 100 and 500 $k\Omega$. No circulator is necessary to isolate input from output, because the two operate at different frequencies, and the filters provide all the isolation required. Note finally that the tripler is provided with an idler circuit, which is a tuned circuit operating at the frequency of $f_{out} - f_{in}$. Idlers are further discussed in conjunction with parametric amplifiers, where the need for them is explained in Section 12-4.1.

Performance, comparison and applications Snap-off varactors multiply by high factors with better efficiency than ordinary varactor chains, and so they are used by preference where possible. Varactors produce higher output powers from about 10 GHz, and step-recovery diodes are not available for frequencies above 20 GHz, while varactors can be used well above 100 GHz. Snap-off devices are suitable for comb generators, whereas the others are not. It has been found that varactor diodes are preferable to step-recovery diodes for broadband frequency multipliers. These are cir-

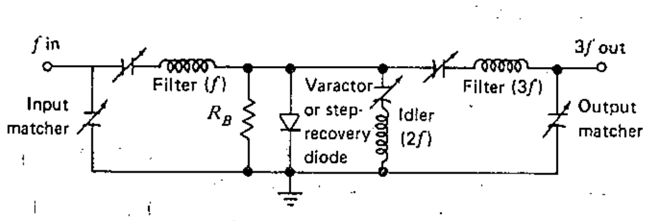

FIGURE 12-12 Diode tripler circuit.

cuits in which the input frequency may occur anywhere within a bandwidth of up to 20 percent, and any such frequency must be multiplied by a given factor.

Step-recovery diodes are available for power outputs in excess of 50 W at 300 MHz, through 10 W at 2 GHz to 1 W at 10 GHz. Multiplication ratios up to 12 are commonly available, and figures as high as 32 have been reported. Efficiency can be in excess of 80 percent for triplers at frequencies up to 1 GHz. With an output frequency of 12 GHz, 5 $\times$ multiplier efficiency drops to 15 percent.

For varactor diodes, the maximum power output ranges from more than 10 W at 2 GHz to about 25 mW at 100 GHz; most varactors at frequencies above 10 GHz are gallium arsenide. Tripler efficiencies range from 70 percent at 2 GHz to just under 40 percent at 36 GHz, and a GaAs varactor doubler efficiency of 54 percent at 60 GHz has also been reported.

For many years, frequency multiplier chains provided the highest microwave powers available from semiconductors, but other developments have overtaken them. At the lower end of the microwave spectrum, GaAs FETs are capable of higher powers, as are Gunn and IMPATT diodes (see following sections) from about 20 to at least 100 GHz. Unless the highest frequency stabilities are required (note that it is the output of a *crystal* oscillator that is multiplied), it is more likely that a transistor Gunn or IMPATT oscillator will be used up to about 100 GHz. One of the current applications of multiplier chains is to provide a low-power signal used to phase-lock a Gunn or IMPATT oscillator.

Varactors are used widely for tuning (see Figure 12-32, for example), for frequency-modulating microwave oscillators, and as the active devices in parametric amplifiers, as will be shown in the next section. They are produced by a mature, well-established manufacturing technique, with consequent good reliability and comparatively low prices.

12-4 PARAMETRIC AMPLIFIERS

Although the use of parametric amplifiers for high-frequency, low-noise amplification is relatively recent, the principles themselves are not at all new. They were first propounded by Lord Rayleigh in the 1880s for mechanical systems, and by R. Hartley in the 1930s for electrical applications. The need for extremely low-noise amplification, as in radiotelescopes, space probe tracking and communications, and tropospheric scatter receivers, spurred the design of useful parametric microwave amplifiers in the late 1950s. The low-noise properties of such amplifiers and the advent of suitable varactor diodes were the greatest stimuli.

12-4.1 Basic Principles

The parametric amplifier uses a device whose reactance is varied in such a manner that amplification results. It is low-noise because no resistance need be involved in the amplifying process. A varactor diode is now always used as the active element. Amplification is obtained when the reactance (capacitive here) is varied electronically in

some predetermined fashion at some frequency *higher* than the frequency of the signal being amplified. The name of the amplifier stems from the fact that capacitance is a *parameter* of a tuned circuit.

Fundamentals To understand the operation of one of the forms of the parametric amplifier, consider an LC circuit oscillating at its natural frequency. If the capacitor plates are physically pulled apart at the instant of time when the voltage between them is at its positive maximum, then work is done on the capacitor since a force must be applied to separate the plates. This work, or energy addition, appears as an increase in the voltage across the capacitor. Since $V = q/C$ and the charge q remains constant, voltage is inversely proportional to capacitance. Since the capacitance has been reduced by the pulling apart of the plates, voltage across them has increased proportionately. The plates are now returned to their initial separation just as the voltage between them passes through zero, which involves no work. As the voltage passes through the negative maximum, the plates are pushed apart, and voltage increases once again. The process is repeated regularly, so that energy is taken from the "pump" source and added to the signal, *at the signal frequency;* amplification will take place if an input circuit and a load are connected. In practice, the capacitance is varied electronically (as could be the inductance). Thus the reactance variation can be made at a much faster rate than by mechanical means, and it is also sinusoidal rather than a square wave.

Comparing the principles of the parametric amplifier with those of more conventional amplifiers we see that the basic difference lies in use of a variable reactance (and an ac power-supply) by the former, and a variable resistance (and a dc power supply) by the latter. As an example, in an ordinary transistor amplifier, changes in base current cause changes in collector current when the collector supply voltage is constant; it may be said that the collector resistance is being changed.

The basic parametric amplifier just described requires the capacitance variation to occur at a *pump* frequency that is exactly twice the resonant frequency of the tuned circuit, and hence twice the signal frequency. It is thus phase-sensitive; this is a property that sometimes limits its usefulness. This mode of operation is called the *degenerate mode,* and it may also be shown that the amplifier is a negative-resistance one (see also Section 12-5.3).

Amplification mechanism The introduction laid down the basis of parametric amplification, and Figure 12-13 illustrates the process graphically. It will be seen that (as outlined) the voltage across the capacitor is increased by pumping at each signal voltage peak. Furthermore, the energy thus given to the circuit is not removed when the plates are restored to their initial position (i.e., when the capacitance of the diode is restored to its original value) because this is done when the voltage across the capacitance is instantaneously zero.

The process of signal buildup is shown in Figure 12-13c. Note that it requires more energy in each successive step to increase the voltage across the capacitance, because the peak charge is greater each time. The capacitor voltage would tend to increase indefinitely, except that the driving power is finite. Thus in practice the buildup progresses until the energy added at each peak equals the maximum energy available from the pump source.

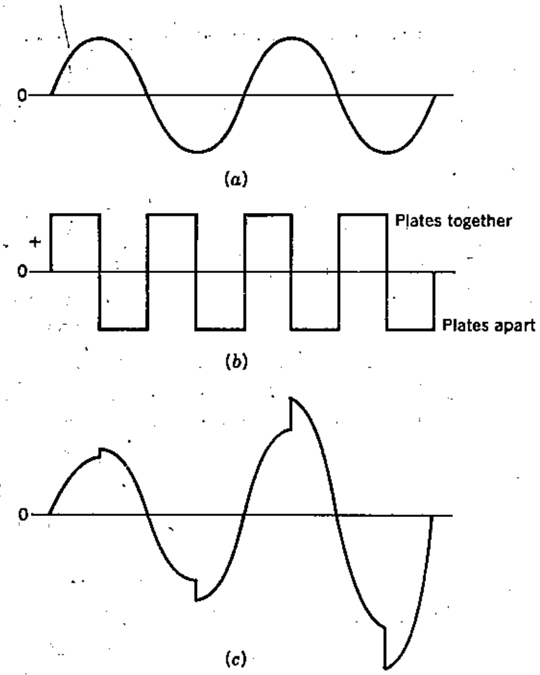

FIGURE 12-13 Parametric amplification with square-wave pumping in degenerate mode. (*a*) Signal input voltage; (*b*) pumping voltage; (*c*) output voltage buildup.

If the pump frequency is other than twice the signal frequency, beating between the two will occur, and a difference signal, called the *idler* frequency, will appear. The amplitude of this idler signal is equal to the amplitude of the output signal, and its presence is an automatic consequence of using a pump frequency such that $f_p \neq 2f_s$. This means that *if the idler signal is suppressed, the amplifier will have no gain.*

Figure 12-14 shows two simple parametric amplifier circuits. In the basic diagram (Figure 12-14*a*) degenerate operation takes place, whereas for Figure 12-14*b* $f_p \neq 2f_s$, and the pumping is called *nondegenerate*. An idler circuit is necessary for amplification to take place, and one is provided. The pump frequency tuned circuit has

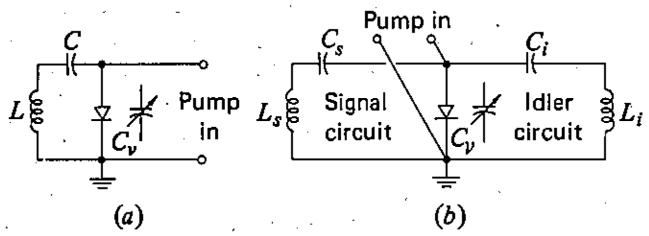

FIGURE 12-14 Basic parametric amplifiers. (*a*) Degenerate; (*b*) nondegenerative, showing idler circuit.

been left out in each case for the sake of simplicity. Note that nothing prevents us from taking the output at the idler frequency, and in fact there are a number of advantages in doing this.

The nondegenerate parametric amplifier, like the degenerate one, produces gain, with the pump source being a net supplier of energy to the tank circuit. This can only be proved mathematically, with the aid of the Manley-Rowe relations. These show that substantial gain is available from this parametric amplifier, in which the pump frequency has no special relationship to the signal frequency (except to be higher, as a general rule). This still holds if sine-wave pumping is used, and it also applies if the output is at the idler frequency.

In the nondegenerate parametric amplifier, the energy taken from the pumping source is transformed into added signal-frequency and idler-frequency energy and divides equally between the two tuned circuits. An amplified output may thus be obtained at either frequency, raising the possibility of frequency conversion *with gain*. In fact, two different types of converters are possible. If the pump frequency is much higher than the signal frequency, then the idler frequency f_i, which is given by $f_i = f_p - f_s$, will be much higher than f_s, and the circuit is called an *up-converter*. If the pump frequency is only slightly higher, f_i will be less than f_s, and a *down-converter*, which is rather similar to the mixer in an ordinary radio receiver, will result. These aspects of parametric amplification will be discussed in detail in the next section.

Note finally that there is no compulsion whatever for the pump frequency to be a multiple of the signal frequency in the nondegenerate amplifier, in fact, it seldom is a multiple in practice.

12-4.2 Amplifier Circuits

The basic types of parametric amplifiers have already been discussed in detail, but several others also exist. They differ from one another in the variable reactance used, the bandwidth required and the output frequency (signal or idler). Various other characteristics of parametric amplifiers must also now be discussed, such as practical circuits, their performance and advantages, and lastly the important noise performance.

Amplifier types When classifying parametric amplifiers, the first thing to decide is the device whose parameter will be varied. This is now always a varactor, whose capacitance is varied, but a variable inductance can also be used. Indeed, the first parametric amplifiers were of this type, using an RF magnetic field to pump a small ferrite disk. Such amplifiers are no longer used, mainly because their noise figures do not compare with those available from varactor amplifiers.

Parametric amplifiers (or *paramps*) may be divided into two main groups; negative-resistance and positive-resistance. The *upper-sideband up-converter* is the only useful member of the second group. Its output is taken at the idler frequency $f_i = f_p + f_s$, and the pump frequency is less than signal frequency. The resulting amplifier has low gain, but a high pumping frequency is not required. This amplifier is most useful at the highest frequencies, for which it was developed.

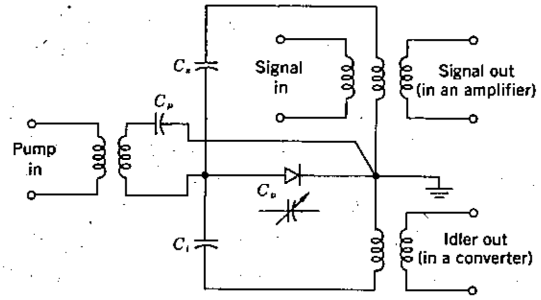

FIGURE 12-15 Parametric amplifier or converter.

Negative-resistance paramps are either straight-out amplifiers ($f_o = f_i$) or lower-sideband converters. If the output is taken at the idler frequency, we have the two-port *lower-sideband up-converter*. Such a circuit is shown in Figure 12-15. The *lower-sideband down-converter* is in the same category. The output is still taken at the idler frequency, but this is now lower than the signal frequency. Both these amplifiers are nondegenerate.

The (straight-out) amplifier may be degenerate or not, depending on whether pump frequency is twice signal frequency. The two types share the disadvantage of being one-port (two-terminal) amplifiers. The nondegenerate amplifier is the one in which the pump frequency is (much) higher than the signal frequency but is quite unrelated to it. The circuit of Figure 12-15 also applies here.

Any paramp can belong to one of two broad classes. First there are narrowband amplifiers using a varactor diode that is part of a tuned circuit. Paramps can be wideband, in which case a number of diodes are used as part of a traveling-wave structure.

Narrowband amplifiers The negative-resistance parametric amplifier is the type almost always used in practice. The most commonly used types are the nondegenerate one-port amplifier and the two-port lower-sideband up-converter, in that order. The circuit of Figure 12-15 could be either type, depending on where the output is taken. The one-port amplifier may suffer from a lack of stability and low gain due mainly to the fact that the output is taken at the input frequency. On the other hand, the pump power is low and so is noise, and the amplifier can be made small, rugged and inexpensive.

Undoubtedly the fundamental drawback of this amplifier, as it stands, is that the input and output terminals are in parallel, as shown in Figure 12-15. This applies to all two-terminal amplifiers. If such an amplifier is followed by a relatively noisy stage such as a mixer, then the noise from the mixer, present at the output of the parametric amplifier, will find its way to the amplifier's input. It will therefore be reamplified, and the noise performance will suffer.

In order to overcome this difficulty, a circulator is used (see Section 10-5.2). The four-terminal version of the Y stripline circulator of Figure 11-39 is particularly suitable. The arrangement is then identical to that of Figure 12-23, with the paramp

replacing the tunnel diode amplifier shown there. The output of the antenna feeds the parametric amplifier, whose output can go only to the mixer. Any noise present at the input of the mixer can be coupled neither to the paramp nor to the antenna; it goes only to the matched termination. The circulator itself can generate some noise, but this may be reduced with proper techniques (such as cooling).

If the output is taken at the idler frequency (in Figure 12-15), a two-port lower-sideband up-converter results, for which a circulator is not required. It has been shown that this type of amplifier is capable of a very low noise figure if f_i/f_s is in excess of about 10. In fact, as this ratio increases, noise figure is lowered, but there are two limitations. The first is the complexity and/or lack of suitably powerful pump sources at millimeter wavelengths, which means that this amplifier is unlikely to be used above X band. The second limitation is the very narrow bandwidth available for minimum noise conditions. The result of all these considerations is that the nondegenerate one-port amplifier (with circulator) is most likely to be used for low-noise narrowband applications.

Traveling-wave diode amplifiers All the parametric amplifiers so far described use cavity or coaxial resonators as tuned circuits. Since such resonators have high *Q's* and therefore narrow bandwidths, parametric amplifiers using them are anything but broadband; the available literature does not describe any such amplifier exceeding a bandwidth of 10 percent. However, it is possible to use traveling-wave structures for parametric amplifiers to provide bandwidths as large as 50 percent of the center frequency, with other properties comparable to those of narrowband amplifiers.

As shown in Figure 12-16, a typical traveling-wave amplifier employs a multistage low-pass filter, consisting of either a transmission line or lumped inductances, with suitably pumped shunt varactor diodes providing the shunt capacitances. The signal and pump frequencies are applied at the input end of the circuit, and the required output is taken from the other end. If the filter is correctly terminated at the desired output frequency, this will not be reflected back to the input, and thus unilateral operation is obtained, even for a negative-resistance amplifier without a circulator. The only real disadvantage is a lower gain than with narrowband amplifiers.

In order to obtain useful amplification, the pump, signal and idler frequencies must all fall within the bandpass of the filter, whereas the sum of the signal and the pump frequencies must fall outside the bandpass. This suggests that the pump frequency must not be very much higher than the signal frequency, or filtering will be difficult. As the wave progresses along the filter (lumped or transmission-line), the signal and idler voltages grow at the expense of the pumping signal. Although this power conversion becomes more complete as the length of the line is increased, the

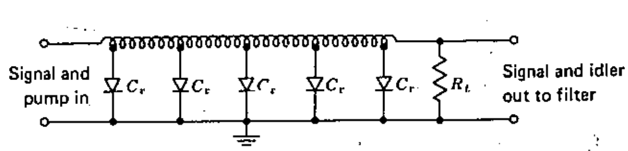

FIGURE 12-16 Basic traveling-wave parametric amplifier.

growth rate reduces. Maximum gain is achieved for a certain optimum length of line (or number of lumped sections), particularly as ohmic losses increase with the length.

Noise—cooling The noise figures of practical parametric amplifiers are extremely low, a very close second only to those of cooled multilevel masers (Section 12-9.2). The reason for such low noise is that the variable transconductance used in the amplifying process is reactive, rather than resistive as in the more orthodox amplifiers. Once noise contributions due to associated circuitry (such as the circulator) have been minimized, the only noise source in the parametric amplifier is the base resistance, sometimes called the *spreading resistance,* of Figure 12-10. This being the case, it seems that cooling the paramp and associated circuitry should have the effect of lowering its noise considerably.

Those paramps that are not operated at room temperature (290 K, or 17°C, is considered standard) may be cooled to about 230 K by using *Peltier thermoelectric cooling*. The next step is to use *cryogenic* cooling with liquid nitrogen (to 77 K) or with liquid helium (4.2 K). Cryogenic apparatus is outside the scope of this book, although one system is shown in outline in Figure 12-28.

It must be emphasized that cooling is used with some parametric amplifiers in an attempt to improve their performance; it is neither compulsory nor always employed. As a matter of fact, although the noise temperature improvement which results from cooling is significant, it is not as great as might be expected. It would appear that the spreading resistance is increased as temperature is lowered, perhaps because of a decrease in the mobility of the varactor's charge carriers. The point is uncertain, however, because measurements at extremely low temperatures are rather difficult to make.

Cryogenic cooling tends to be bulky and expensive, and consequently the current trend is away from cryogenically cooled amplifiers, except for the most exacting applications, as in radiotelescopes, some satellite earth stations, and space communication terminals. Thus applications requiring very good but not critical noise figures, including portable earth stations, are likely to use Peltier cooled or uncooled paramps. The other current design feature is the use of solid-state (especially Gunn) oscillators for pumps, although a lot of existing parametric amplifiers still use klystrons or even varactor chains.

The Ferranti parametric amplifier of Figure 12-17 is uncooled. In the figure, terminal 1 is the RF input, 2 is the pump input, 3 the RF output and 4 the connection point for temperature stabilizing equipment. The circulator is behind the paramp box, which measures approximately 70 × 45 × 25 mm. The Gunn diode pump source is located next to the circulator, together with its cavity and thermal stabilizer. The panel has its own dc supply and simply plugs into the mains. Two diodes are connected back-to-back in a coaxial circuit, using a system pioneered by the firm. The varactors in the paramp are high-quality, high-frequency GaAs ones.

Performance comparisons There are so many different types of parametric amplifiers and temperatures at which they may be used that tabular comparison is considered the most convenient. Accordingly, Table 12-1 compares a number of typical paramps; note the degradation in noise figure with increased temperature and/or operating fre-

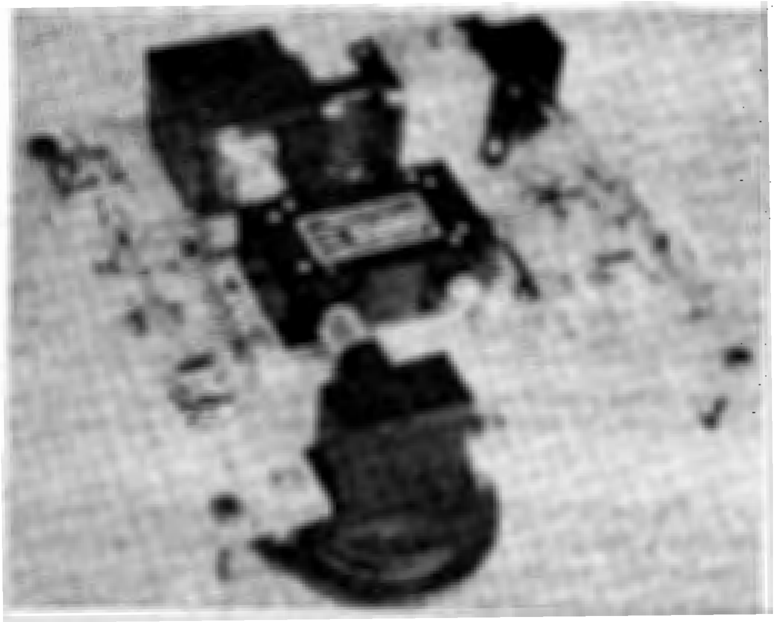

FIGURE 12-17 Uncooled 5-GHz parametric amplifier with ancillary equipment. ***(Courtesy of Ferranti Ltd., Solid State Microwave Group.)***

quency. Note also the lower bandwidth of converters as compared with nondegenerate one-port amplifiers, while the traveling-wave amplifier has by far the greatest percentage bandwidth.

The comparison in Table 12-2 is between paramps and other low-noise amplifiers. Note that the best, rather than typical, performances are included in Table 12-2.

TABLE 12-1 Performance Comparison of Various Parametric Amplifier Types

AMPLIFIER TYPE	WORKING TEMPERATURE, K	f_{in}, GHz	f_p, GHz	f_{out}, GHz	POWER GAIN, dB	BAND WIDTH, MHz	NOISE FIGURE, dB	NOISE TEMPERATURE, K
Degenerate*	4.2	6.00	12.0	6.00	14	10	0.3	21
Degenerate*	290	5.85	11.7	5.85	18	8	3.0	300
Nondegenerate*	4.2	4.2	23.0	4.2	22	40	0.2	14
Nondegenerate*	77	4.1	23.0	4.1	20	60	0.6	45
Nondegenerate*	290	3.95	61.0	3.95	60	500	1.0	80
(Not known)*	235	3.95	?	3.95	60	500	0.75	55
Nondegenerate*	290	60.0	105.0	60.0	14	670	6.0	865
LSB up-converter	290	0.9	26.5	25.6	16	2.5	1.0	80
USB up-converter	77	1.0	20.0	21.0	10	0.1	0.4	29
Traveling-wave	290	3.4	8.5	3.4	10	720	3.5	370

*All these amplifiers are one-port and hence require circulators.

TABLE 12-2 Comparison of Various Low-Noise Amplifiers*

TYPE	f_{out}, GHz	POWER GAIN, dB	BAND WIDTH, MHz	NOISE TEMPERATURE, K	COOLING
Parametric amplifier	4.00	19	40	8	Very helpful
Traveling-wave paramp	4.10	12	500	16	
Three-level ruby maser	8.00	10	5	6	Compulsory (with
Traveling-wave maser	5.80	20	25	11	liquid helium)
Tunnel-diode amplifier	4.00	30	75	400	Helps (but de-
Tunnel-diode amplifier	3.00	10	2,000	500	stroys simplicity)
GaAs FET amplifier	3.00	32	2,000	200	As above
Low-noise TWT	3.00	25	2,000	600	Not practicable

*The figures shown are for the best available commercial amplifiers, of which the paramps and masers are cooled down to 4.2 K. Typical noise temperatures for mixers, which may be used instead, are approximately 700 K.

Parametric amplifiers find use in microwave receivers which require extremely low-noise temperatures. At the lowest point, in radiotelescopes and satellite and space probe tracking stations, they compete with masers. They are used in earth stations, sometimes in communications satellites and, increasingly, in radar receivers.

12-5 TUNNEL DIODES AND NEGATIVE-RESISTANCE AMPLIFIERS

The tunnel, or Esaki, diode is a thin-junction diode which, under low forward-bias conditions, exhibits negative resistance. This makes the tunnel diode, invented in the late 1950s, useful for oscillation or amplification. Because of the thin junction and short transit time, it lends itself well to microwave applications.

12-5.1 Principles of Tunnel Diodes

The equivalent circuit of the tunnel diode, when biased in the negative-resistance region, is shown in Figure 12-18. At all except the highest frequencies, the series resistance and inductance can be ignored. The resulting diode equivalent circuit is thus reduced to the parallel combination of the junction capacitance C_j and the negative resistance $-R$. Typical values of the circuit components of Figure 12-18 are $r_s = 6\ \Omega$, $L_s = 0.1$ nH, $C_j = 0.6$ pF and $R = -75\ \Omega$.

The tunnel diode was the first, and for several years the only, solid-state device which merely required the application of a small dc voltage for negative resistance to

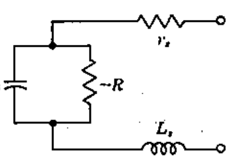

FIGURE 12-18 Tunnel-diode equivalent circuit.

manifest itself. However, after the initial exclamations of joy at the remarkable invention died down, tunnel diode oscillators were found to be unstable in their frequency of operation and were all but discarded. The reasons for the instability have subsequently been found and cured, and so tunnel diodes are again in use, but now mainly for amplifiers.

The junction capacitance of the tunnel diode is highly dependent on the bias voltage and temperature. Connecting a tuned circuit directly across it will undoubtedly yield an unstable oscillator, particularly since the effective Q of the circuit is relatively low. However, if a high-Q cavity is *loosely coupled* to the diode, a highly stable oscillator is obtained, with a relative independence of temperature, bias voltage or diode parameter variation.

Description of behavior The tunnel diode is a semiconductor *p-n* junction diode. It differs from the usual rectifier-type diodes in that the semiconductor materials are very heavily doped, perhaps as much as 1000 times more than in ordinary diodes. This heavy doping results in a junction which has a depletion layer that (with a typical thickness of 0.01 μm) is so thin as to prevent *tunneling* to occur. In addition, the thinness of the junction allows microwave operation of the diode because it considerably shortens the time taken by the carriers to cross the junction. A current-voltage characteristic for a typical germanium tunnel diode is shown in Figure 12-19. It is seen that at first forward current rises sharply as voltage is applied, where it would have risen slowly for an ordinary diode (whose characteristic is shown for comparison). Also, reverse current is much larger for comparable back bias than in other diodes, owing to the thinness of the junction.

The interesting portion of the characteristic begins at the point A on the curve of Figure 12-19; this is thc *voltage peak*. As the forward bias is increased past this point, the forward current drops and continues to drop until point B is reached; this is the *valley voltage*. At B the current starts to increase once again and does so very rapidly as bias is increased further. From this point the characteristic resembles that of an ordi-

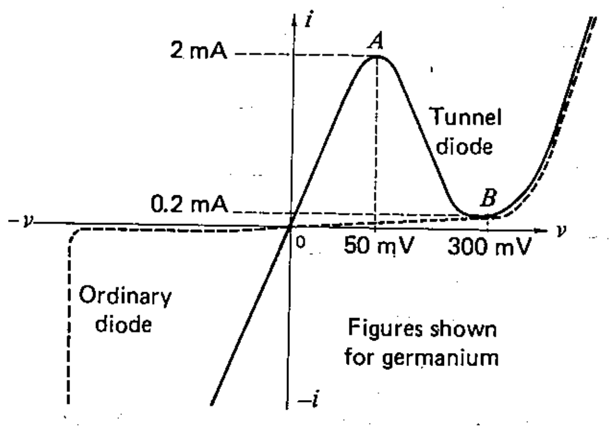

FIGURE 12-19 Tunnel-diode voltage-current characteristic.

nary diode. Apart from the voltage peak and valley, the other two parameters normally used to specify the diode behavior are the peak current and the peak-to-valley current ratio, which here are 2 mA and 10, respectively, as shown.

The diode voltage-current characteristic illustrates two important properties of the tunnel diode. First it shows that the diode exhibits *dynamic negative resistance* between A and B and is therefore useful for oscillator (and amplifier) applications. Second since this negative resistance occurs when both the applied voltage and the resulting current are low, the tunnel diode is a relatively low-power device. A quick calculation shows that in order to stay within the negative-resistance region, the voltage variation must be restricted to $300 - 50 = 250$ mV (peak-to-peak) = 88.4 mV rms, whereas the current range is similarly 1.8 mA (peak-to-peak) = 0.63 mA. The load power is very roughly $88.4 \times 0.635 = 56\ \mu$W. Other factors have been neglected, but the figure is of the right order.

Diode theory Unless energy is imparted to electrons from some external source, the energy possessed by the electrons on the n side of the junction is insufficient to permit them to *climb over* the junction barrier to reach the p side. *Quantum mechanics* shows that there is a small but finite probability that an electron which has insufficient energy to climb the barrier can, nevertheless, find itself on the other side of it if this barrier is thin enough, without any loss of energy on the part of the electron. This is the tunneling phenomenon which is responsible for the behavior of the diode over the region of interest.

Figure 12-20 shows energy-level diagrams for the tunnel diode for three interesting bias levels. The cross-hatched regions represent energy states in the conduction band occupied by electrons, whereas the shaded areas show the energy states occupied by electrons in the valence bands. The levels to which energy states are occupied by

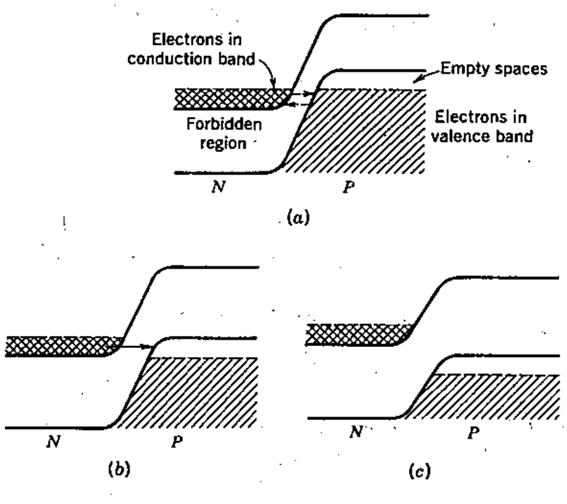

FIGURE 12-20 Energy-level diagrams for tunnel-diode junction at (*a*) zero bias voltage; (*b*) peak voltage; (*c*) valley voltage. ***(Courtesy of RCA.)***

electrons on either side of the junction are shown by dotted lines. When the bias voltage is zero, these lines are at the same height. Electrons can now tunnel from one side of the junction to the other because of its thinness, but the tunneling currents in the two directions are the same. No effective overall current flows. This is shown in Figure 12-20*a*.

When a small forward bias is applied to the junction, the energy level of the *p* side is lowered (as compared with the *n* side). As shown in Figure 12-20*b*, electrons are able to tunnel through from the *n* side. This is possible because the electrons in the conduction band there find themselves opposite vacant states on the *p* side. Tunneling in the other direction is not possible, because the valence-band electrons on the *p* side are now opposite the forbidden energy gap on the *n* side. This gap, shown here at its maximum, represents the peak of the diode characteristic.

When the forward bias is raised beyond this point, tunneling will decrease, as may be seen with the aid of Figure 12-20*c*. The energy level on the *p* side is now depressed further, with the result that fewer *n*-side free electrons are opposite unoccupied *p*-side energy levels. As the bias is raised, forward current drops; this corresponds to the negative-resistance region of the diode characteristic. As Figure 12-20*c* shows, a forward bias is reached at which there are no conduction-band electrons opposite valence-band vacant states, and tunneling stops altogether. The point at which this happens is the valley of Figure 12-19, to which the energy-level diagram of Figure 12-20*c* corresponds. When forward voltage is increased even further, "normal" forward current flows and increases, as with ordinary rectifier diodes.

It is thus seen that the curious phenomenon in tunnel diodes is not only the negative-resistance region but also the forward current peak that precedes it. As a result of tunneling across the narrow junction, forward current flows initially in much greater quantities than in a rectifier diode. As the forward bias is raised, tunneling becomes more difficult, the tunneling current is reduced and the negative-resistance region results. As the increase in forward voltage continues, tunneling stops completely, and the normal operation takes over. The valley is the point at which this "return to normalcy" begins.

Materials and construction Although tunnel diodes could be made from any semiconductor material, initially germanium and then gallium antimonide and gallium arsenide have been preferred in practice. All have small forbidden energy gaps and high ion mobilities, which are characteristics leading to good high-frequency or high-speed operation. These materials are preferable to silicon and other semiconductors in this regard.

As the cross section of Figure 12-21 shows, the construction of a tunnel diode is remarkably simple. This is yet another advantage of the device, particularly since the fabrication is also simple. A very small tin dot, about 50 μm in diameter, is soldered or alloyed to a heavily doped pellet (about 0.5 mm square) of *n*-type Ge, GaSb or GaAs. The pellet is then soldered to a Kovar pedestal, used for heat dissipation, which forms the anode contact. The cathode contact is also Kovar, being connected to the tin dot via a mesh screen used to reduce inductance. The diode has a ceramic body and a hermetically sealing lid on top. Note the tiny dimensions of the pill package.

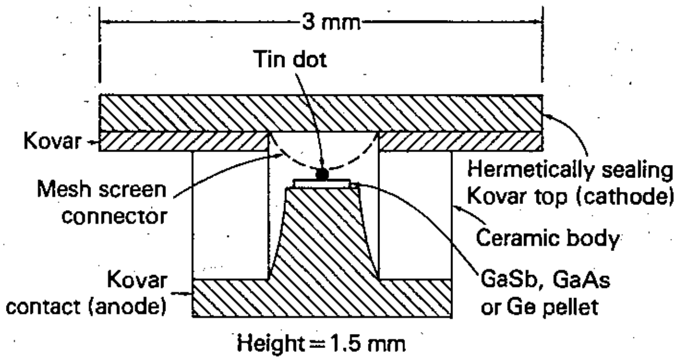

FIGURE 12-21 Construction of typical tunnel diode.

12-5.2 Negative-Resistance Amplifiers

The classical application of the tunnel diode was in microwave oscillators, especially after it was realized that the secret of stable oscillations lay in loosely coupling the diode to its tuned circuit. Other semiconductor devices have subsequently appeared, producing far more microwave power than the tunnel diode ever could. The tunnel diode has been superseded in some of its traditional oscillator applications. It is important to realize that the tunnel diode is a fully fledged active device, like the transistor, so that amplification may be performed with it. *It will now be used as a vehicle to introduce negative-resistance amplifiers in general.* These are common at microwaves, and indeed negative-resistance parametric amplifiers have already been met.

Theory of negative-resistance amplifiers It can be shown that a circuit incorporating a negative resistance is capable of significant power gain. This is obvious, since negative-resistance oscillators are able to oscillate, it is clear that the negative resistance must be making up all the circuit losses. It feeds power into the circuit, which dissipates some and puts out the rest. This is similar to the feedback oscillator situation, in which βA must at least equal unity, and therefore gain certainly exists. The proof for the tunnel diode now follows, but it is really independent of the particular device used to provide the negative resistance.

Consider the basic negative-resistance amplifier of Figure 12-22. It consists of an input current source i_s, together with the source conductance g_s, connected to a negative conductance $-g$. Across this the load conductance g_L is also connected. The current source and parallel circuit are used for ease of proof. If the frequency is not so high that r_s and L_s of the tunnel-diode equivalent circuit must be taken into account,

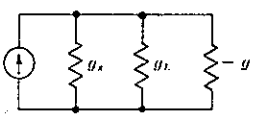

FIGURE 12-22 Basic negative-resistance amplifier.

and if the junction capacitance C_j is tuned out, the $-g$ is a suitable representation of the tunnel diode. *In the absence of the diode,* the maximum power available from the generator will be when $g_L = g_s$, i.e.,

$$P_{\max} = \frac{i_s^2}{4g_s} \tag{12-4}$$

With the diode present, the load voltage is

$$v_L = \frac{i_s}{g_s - g + g_L} \tag{12-5}$$

The power delivered to the load is

$$P_L = v_L^2 g_L = \frac{g_L i_s^2}{(g_s - g + g_L)^2} \tag{12-6}$$

If the presence of the diode has permitted power gain, the ratio of Equation (12-5) to Equation (12-4) is greater than unity. Then

$$\begin{aligned} A_P &= \frac{P_L}{P_{\max}} = \frac{i_s^2 g_L/(g_s - g + g_L)^2}{i_s^2/4g_s} \\ &= \frac{4g_s g_L}{(g_s - g + g_L)^2} \end{aligned} \tag{12-7}$$

For maximum power transfer, the load and generator conductances are made equal as before. With this new condition we have

$$\begin{aligned} A_P &= \frac{4g_L^2}{(2g_L - g)^2} \\ &= \frac{4g_L^2}{4g_L^2 - 4g_L g + g^2} \\ &= \frac{4g_L^2}{4g_L^2 + g(g - 4g_L)} \end{aligned} \tag{12-8}$$

Equation (12-8) can obviously be greater than 1, provided that the second term in its denominator is negative, i.e., *provided that* $4g_L$ *is greater than* g. If this applies, A_P exceeds unity, real power gain is available, and the circuit may be used as an amplifier. Care must be taken to ensure that the denominator of Equation (12-8) is not reduced to zero, which would happen for a value of g such that the last term of Equation (12-8) is equal to -1. Simple algebra shows that this would occur when

$$g = 2g_L \qquad (\text{if } g_L = g_s \text{ as before}) \tag{12-9}$$

It is seen that an amplifier containing a negative resistance is capable not only of power gain but also of *infinite gain* (and therefore oscillation). This occurs when Equation (12-9) holds, and it gives the lower limit for the value of g, and hence the upper limit for the value of the negative resistance. (Note that the lower limit of the negative resistance is governed by the requirement that $4g_L$ must be greater than g.) We

have thus proved that the negative-resistance amplifier is capable of power gain if the negative resistance has a value between the limits just described. If it strays outside these limits, either Equation (12-8) exceeds unity, and therefore power gain is less than 1, or else it becomes negative, and oscillations take place.

Tunnel-diode amplifier theory For frequencies below self-resonance, Equation (12-7) must be enlarged to include the junction capacitance of the diode. This capacitance is tuned out in an amplifier, but including it yields a useful result. Therefore

$$A_P = \frac{4g_s g_L}{(g_s + g_L - g + j\omega C_j)^2} \tag{12-10}$$

This, in turn, gives a resistive cutoff frequency, or figure of merit, for such a diode, which corresponds to the frequency at which the magnitude of ωC_j equals the magnitude of $-g$. Past this frequency, the negative resistance of the tunnel diode disappears. This frequency is given by

$$g = \omega_r C_j$$

$$R = \frac{1}{\omega_r C_j}$$

$$\omega_r = \frac{1}{RC_j}$$

$$f_r = \frac{1}{2\pi RC_j} \tag{12-11}$$

The series diode loss resistance r_s of Figure 12-18 has been neglected in this derivation, because it is much smaller than the negative resistance (generally being no more than one-tenth of the negative resistance) and thus its effect is very small. An alternative interpretation of Equation (12-11) is that it represents the gain-bandwidth product of a tunnel-diode amplifier.

12-5.3 Tunnel-Diode Applications

In all its applications, the tunnel diode should be loosely coupled to its tuned circuit. With lumped components, this is done by means of a capacitive divider, with the diode connected to a tapping point, while the divider is across the tuned circuit itself. In a cavity, the diode is placed at a point of significant, but not maximum, coupling. The other point of significance is the application of dc bias. This must be connected to the diode without interfering with the tuned circuit. The simplest way of doing this is with a filter, as shown in Figure 12-23. Basically, this filter prevents the diode from being short-circuited by the supply source, while ensuring that no positive resistance is added to interfere with the negative resistance of the diode. Also, the addition of capacitance across the diode is avoided. Care must be taken to ensure that the bias inductance does not introduce spurious frequencies in the bandpass.

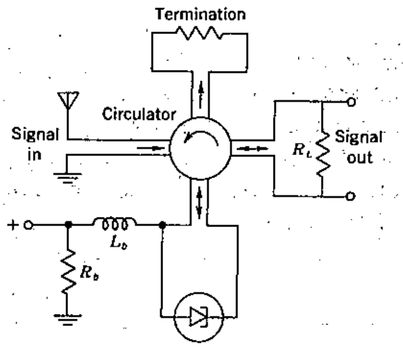

FIGURE 12-23 Tunnel-diode amplifier with circulator. *(Based on a figure from "Tunnel Diodes," by courtesy of RCA.)*

Amplifiers As shown in Figure 12-23, the tunnel-diode amplifier (TDA), like the parametric amplifier, requires a circulator to separate the input from the output. Their layouts are very similar, with the very significant difference that no pump source is required for the TDA.

Tables 12-1 and 12-2 show a number of low-noise microwave amplifier performance figures, including those of tunnel-diode amplifiers. It is seen that the tunnel diode is a low-noise device. The twin reasons for this are the low value of the parasitic resistance r_s (producing low thermal noise) and the low operating current (producing low shot noise). In such low-noise company, TDAs are as broadband as any, are very small and simple and have output levels on a par with paramps and masers. The available gains are high, and operating frequencies in excess of 50 GHz have been reported.

Amplifier applications Tunnel-diode amplifiers may be used throughout the microwave range as moderate-to-low-noise preamplifiers in all kinds of receivers. GaAs FET amplifiers are more likely to be used in current equipment up to 18 GHz. Large bandwidths and high gains are available from multistage amplifiers, the circuits and power requirements are very simple (typically a few milliamperes at 10 V-dc), and noise figures below 5 dB are possible well above X band. It is worth noting that TDAs are immune to the ambient radiation encountered in interplanetary space, and so are practicable for space work.

Other applications Tunnel diodes are *diodes* that may be used as mixers. Being also capable of active oscillation, they may be used as self-excited mixers, in a manner similar to the transistor mixer of Section 6-2.2. Being high-speed devices, tunnel diodes also lend themselves to high-speed switching and logic operations, as flip-flops and gates. They are used as low-power oscillators up to about 100 GHz, because of their simplicity, frequency stability and immunity to radiation.

12-6 GUNN EFFECT AND DIODES

It was being prophetically said in the early 1960s that as long as the important properties of semiconductors depended on junctions—and these had to be made thinner and thinner as frequency was raised—then high powers from such semiconductor devices were simply not possible at microwave frequencies, certainly not above X band. Fortunately, since then other devices have been invented or adapted which do not depend on junction behavior and which are capable of producing adequate microwave powers. One of these classes of devices depends on controlled avalanche and takes into account the transit time in its operation. This *Avalanche and Transit Time* family is covered in Section 12-7. The other class of devices exhibits microwave power properties that depend on the behavior of *bulk* semiconductors, rather than junctions. The *Gunn effect* is the main representative of this class of devices and is now discussed.

12-6.1 Gunn Effect

In 1963, Gunn discovered the *transferred electron* effect which now bears his name. This effect is instrumental in the generation of microwave oscillations in bulk semiconductor materials. The effect was found by Gunn to be exhibited by *gallium arsenide* and *indium phosphide,* but *cadmium telluride* and *indium arsenide* have also subsequently been found to possess it. Gunn's discovery was a breakthrough of great importance. It marked the first instance of useful semiconductor device operation depending on the *bulk properties* of a material.

Introduction If a relatively small dc voltage is placed across a thin slice of gallium arsenide, such as the one shown in Figure 12-24, then negative resistance will manifest itself under certain conditions. These consist merely of ensuring that the voltage gradient across the slice is in excess of about 3300 V/cm. Oscillations will then occur if the slice is connected to a suitably tuned circuit. It is seen that the voltage gradient across the slice of GaAs is very high. The electron velocity is also high, so that oscillations will occur at microwave frequencies.

It must be reiterated that the Gunn effect is a *bulk* property of semiconductors and does not depend, as do other semiconductor effects, on either junction or contact

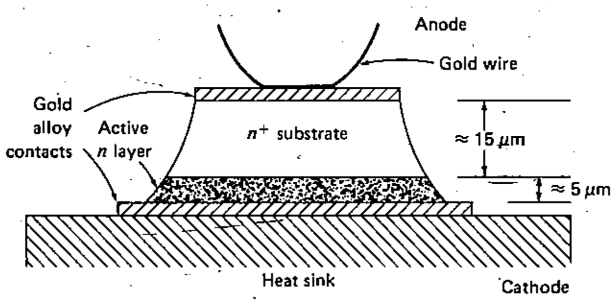

FIGURE 12-24 Epitaxial GaAs Gunn slice.

properties. As established painstakingly by Gunn, the effect is independent of *total* voltage or current and is not affected by magnetic fields or different types of contacts. It occurs in *n-type* materials *only*, so that it must be associated with electrons rather than holes. Having determined that the voltage required was proportional to the sample length, the inventor concluded that the electric field, in volts per centimeter, was the factor determining the presence or absence of oscillations. He also found that a threshold value of 3.3 kV/cm must be exceeded if oscillations are to take place. He found that the frequency of the oscillations produced corresponded closely to the time that electrons would take to traverse such a slice of n-type material as a result of the voltage applied. This suggests that a bunch of electrons, here called a *domain*, is formed somehow, occurs once per cycle and arrives at the positive end of the slice to excite oscillations in the associated tuned circuit.

Negative resistance Although the device itself is very simple, its operation (as might be suspected) is not quite so simple. Gallium arsenide is one of a fairly small number of semiconductor materials which, in an n-doped sample, have an empty energy band higher in energy than the highest filled (or partly filled) band. The size of the forbidden gap between these two is relatively small. This does not apply to some other semiconductor materials, such as silicon and germanium. The situation for gallium arsenide is illustrated in Figure 12-25, in which the highest levels shown also have the highest energies.

When a voltage is applied across a slice of GaAs which is doped so as to have excess electrons (i.e., n-type), these electrons flow as a current toward the positive end of the slice. The greater the potential across the slice, the higher the velocity with which the electrons move toward the positive end, and therefore the greater the current. The device is behaving as a normal positive resistance. In other diodes, the component of velocity toward the positive end, imparted to the electrons by the applied voltage, is quite small compared to the random thermal velocity that these electrons possess. In this case, so much energy is imparted to the electrons by the extremely high voltage gradient that instead of traveling faster and therefore constituting a larger current, their flow actually slows down. This is because such electrons have acquired enough energy to be transferred to the higher energy band, which is normally empty, as shown in Figure 12-25. This gives rise to the name *transferred-electron* effect, which is often

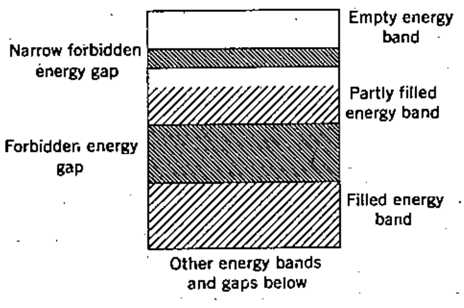

FIGURE 12-25 Important energy levels in gallium arsenide.

given to this phenomenon. *Electrons have been transferred from the conduction band to a higher energy band in which they are much less mobile, and the current has been reduced as a result of a voltage rise.* Note that, in a sense, gallium arsenide is a member of a group of unusual semiconductor substances. In a lot of others, the energy required for this transfer of electrons would be so high, because of a higher forbidden energy gap, that the complete crystal structure might be distorted or even destroyed by the high potential gradient before any transfer of electrons could take place.

It is seen that as the applied voltage rises past the *threshold negative-resistance value,* current falls, and the classical case of negative resistance is exhibited. Eventually the voltage across the slice becomes sufficient to remove electrons from the higher-energy, lower-mobility band, so that current will increase with voltage once again. The voltage-current characteristic of such a slice of gallium arsenide is seen to be very similar to that of a tunnel diode, but for vastly different reasons.

Gunn domains It was stated in the preceding section that the oscillations observed in the initial GaAs slice were compatible with the formation and transit time of electron bunches. It follows, therefore, that the negative resistance just described is not the only effect taking place. The other phenomenon is the formation of *domains,* the reasons for which may now be considered.

It is reasonable to expect that the density of the doping material is not completely uniform throughout our sample of gallium arsenide. Hence it is entirely possible that there will be a region, perhaps somewhere near the negative end, where the impurity concentration is less than average. In such an area there are fewer free electrons than in other areas, and therefore this region is less conductive than the others. As a result of this, there will be a greater than average potential across it. Thus, as the total applied voltage is increased, this region will be the first to have a voltage across it large enough to induce transfer of electrons to the higher energy band. In fact, such a region will have become a *negative-resistance domain.*

A domain like this is obviously unstable. Electrons are being taken out of circulation at a fast rate within it, the ones behind bunch up and the ones in front travel forward rapidly. In fact, the whole domain moves across the slice toward the positive end with the same average velocity as the electrons before and after it, about 10^7 cm/s in practice. Note that such a domain is self-perpetuating. As soon as some electrons in a region have been transferred to the less conductive energy band, fewer free electrons are left behind. Thus this particular region becomes less conductive, and therefore the potential gradient across it increases. The domain is quite capable of traveling and may be thought of as a low-conductivity, high-electron-transfer region, corresponding to a negative pulse of voltage. When it arrives at the positive end of the slice, a pulse is received by the associated tank circuit and shocks it into oscillations. It is actually this arrival of pulses at the anode, rather than the negative resistance proper, which is responsible for oscillations in Gunn diodes. (The term *diode* is a misnomer for Gunn devices since there is no junction, nor is rectification involved. The device is called a *diode* because it has two terminals, and the name is also convenient because it allows the use of *anode* for the "positive end of the slice.")

With the usual applied voltages, once a domain forms, insufficient potential is left across the rest of the slice to permit another domain to form. This assumes that the

sample is fairly short; otherwise the situation can become very complex, with the possibility that other domains may form. The domain described is sometimes called a *dipole domain*. An *accumulation domain* may also occur (particularly in a longer sample), where a more highly doped region is involved, and a current accumulation travels toward the anode. When the domain in a short sample arrives at the anode, there is once again sufficient potential to permit the formation of another domain somewhere near the cathode. It is seen that only one domain, or pulse, is formed per cycle of RF oscillations, and so energy is received by the tank circuit in correct phase to permit the oscillations to continue.

Miscellaneous considerations The following brief notes and observations are now made:

1. The Gunn diode/oscillator has received various names. Although *Gunn* (diode or oscillator) will always be used here, students should be aware of names such as *transferred-electron devices* (or *TED*) and *transferred-electron oscillator* (or *TEO*), which are also frequently used.
2. The Gunn effect is eminently viable, and Gunn diodes are very widely used.
3. Although only gallium arsenide diodes have been considered, indium phosphide (InP) Gunn diodes are becoming widely used, especially at the highest frequencies. InP has properties quite similar to those of GaAs, while also offering the advantages of a higher peak-to-valley ratio in its negative resistance characteristic and lower noise.
4. Students must surely be wondering by now why such "oddball" operating principles are used by the majority of solid-state microwave devices—and they have not even encountered the IMPATT diode as yet! The only explanation the author can put forward is that all the simple devices got themselves invented a long time ago, and *they* didn't work at microwave frequencies!

12-6.2 Gunn Diodes and Applications

Gunn diodes A practical Gunn diode consists of a slice like the one shown in Figure 12-24, sometimes with a buffer layer between the active layer and the substrate, mounted in any of a number of packages, depending on the manufacturer, the frequency and the power level. Encapsulation identical to that shown for varactor diodes in Figure 12-9 is common. The power that must be dissipated is quite comparable.

Gunn diodes are grown epitaxially out of GaAs or InP doped with silicon, tellurium or selenium. The substrate, used here as an ohmic contact, is highly doped for good conductivity, while the thin active layer is less heavily doped. The gold alloy contacts are electrodeposited and used for good ohmic contact and heat transfer for subsequent dissipation. Diodes have been made with active layers varying in thickness from 40 to about 1 μm at the highest frequencies. The actual structure is normally square, and so far GaAs diodes predominate commercially.

Diode performance As a good approximation, the equivalent circuit of a GaAs X-band Gunn diode consists of a negative resistance of about 100 ohms (100 Ω) in

parallel with a capacitance of about 0.6 pF. Such a commercial diode will require a 9-V dc bias, and, with an operating current of 950 mA, the dissipation in its (cathode) heat sink will be 8.55 W. Given that the output (anywhere in the range 8 to 12.4 GHz) is 300 mW, the efficiency is seen to be 3.5 percent. A higher-frequency Gunn diode, operating over the range of 26.5 to 40 GHz, might produce an output of 250 mW with an efficiency of 2.5 percent.

Overall, GaAs Gunn diodes are available commercially for frequencies from 4 GHz (1 to 2 W CW maximum) to about 100 GHz (50 mW CW maximum). Over that range, the maximum claimed efficiencies drop from 20 to about 1 percent, but for most commercial diodes 2.5 to 5 percent is normal. InP diodes, not yet as advanced commercially, have a performance that ranges from 500 mW CW at 45 GHz (efficiency of 6 percent) to 100 mW CW at 90 GHz (efficiency of 4.5 percent); higher powers and operating frequencies are expected. Other options available include two or more diodes in one oscillator package for higher CW outputs, and diodes for pulsed outputs. In the latter case, commercial diodes produce up to a few dozen watts pulsed, with 1 percent duty cycles and efficiencies somewhat better than for CW diodes.

Gunn oscillators Since the Gunn diode consists basically of a negative resistance, all that is required in principle to make it into an oscillator is an inductance to tune out the capacitance, and a shunt load resistance not greater than the negative resistance. This has already been discussed in conjunction with the tunnel diode. In practice, a coaxial cavity operating in the TEM mode has been found the most convenient for fixed frequency (but with some mechanical tuning) operation. A typical coaxial Gunn oscillator is shown in Figure 12-26. If some electrical tuning is required as well, a varactor

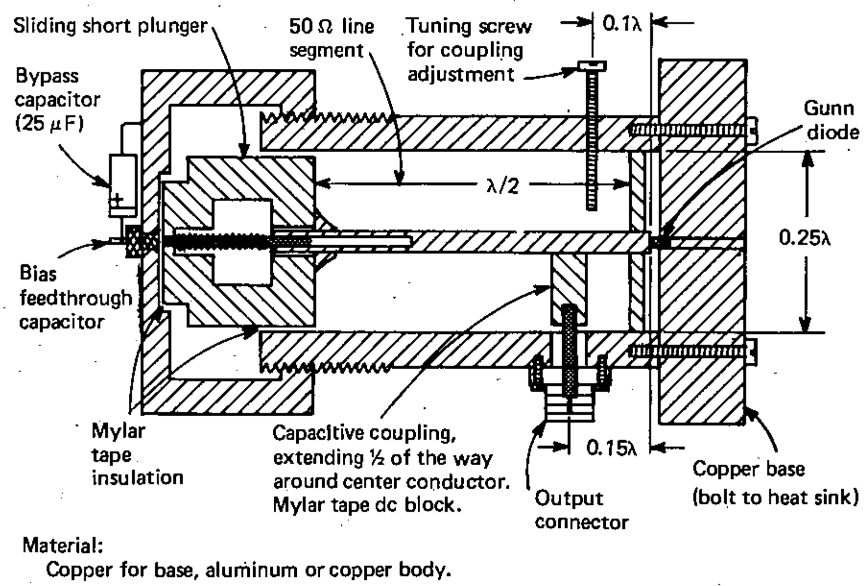

FIGURE 12-26 Cross section of typical Gunn coaxial oscillator cavity. ***(Courtesy of Microwave Associates International.)***

may be placed in the cavity, at the opposite end to the Gunn diode. The dimensions shown in Figure 12-26 are selected to provide suitable diode mounting and dissipation, as well as freedom from spurious mode oscillations.

YIG-tuned (see Section 10-5.2) Gunn VCOs are available for instrument applications, featuring frequency ranges as large as 2 octaves, much greater than is possible with varactors. A selection of such oscillators is shown in Figure 12-27. The 300-g, 50 × 50 mm package contains a Gunn slice on a heat sink, and a cavity with a small YIG sphere. There is a heater for the YIG sphere, to keep it at a constant temperature, and a coil for altering the magnetic field. The instantaneous frequency of oscillation is governed by the cavity frequency, which in turn depends on the YIG sphere and the magnetic field by which it is surrounded. It is the Gunn diode, rather than the tuning mechanism, that determines the frequency limits. When the frequency of the resonator is changed, the diode itself responds by generating its domain at a distance from the anode such that the transit time of the domain corresponds to a cycle of oscillations. As frequency is raised, the formation point of the domain moves closer to the anode. The oscillations eventually stop when this point is more than halfway across the slice. Avantek oscillators of the type shown in Figure 12-27 cover the range from 1 to 12.4 GHz, with typically octave tuning (or sweeping) ranges for each. Frequency modulation is also possible, via the terminals provided, and in all very rapid frequency changes can be made. Such VCOs are designed as backward-wave oscillator replacements, certainly at the lower end of the BWO's operating spectrum. Typical power outputs range up to 50 mW, and total power consumption may be 5 W, including power for the YIG sphere.

Finally, it should be mentioned that the noise performance of Gunn oscillators is quite acceptable. Spurious AM noise is on par with that of the klystron (which itself is very good), while spurious FM noise is worse, but not too high for normal applications. Injection locking with a low-amplitude, high-stability signal helps to reduce FM noise quite significantly.

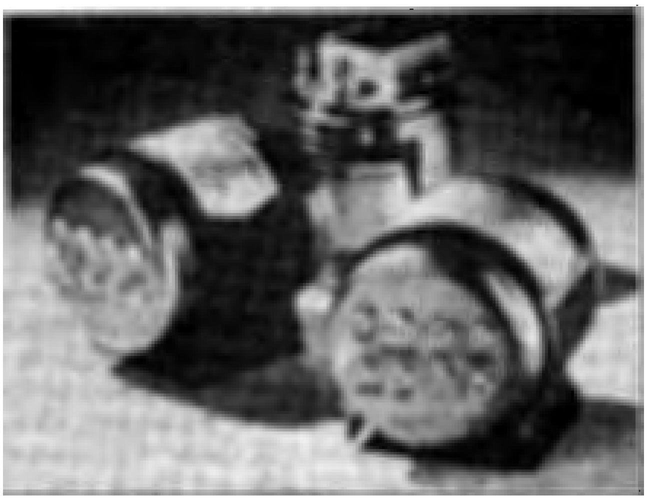

FIGURE 12-27 YIG-tuned voltage-controlled oscillators. ***(Courtesy of Avantek, Inc.)***

Gunn diode amplifiers As was shown in connection with the tunnel diode, a device exhibiting negative resistance may be used as an amplifier, and of course the Gunn diode qualifies in this respect. However, Gunn diode amplifiers are not used nearly as much as Gunn oscillators. The reasons are many. On the one hand, Gunn diode amplifiers cannot compete for power output and low noise with GaAs FET amplifiers at frequencies below about 30 GHz, and at higher frequencies they cannot compete with the power output or efficiency of electron tube or IMPATT (see next section) amplifiers. Accordingly, the niche which is left for them is as low- to medium-power medium-noise amplifiers in the 30- to 100-GHz frequency range. Over that range, they are capable of amplifying with noise figures of the order of 20 to 30 dB, relatively low efficiency and power gain per stage, and an output power that is perhaps two to four times as high would be expected from a single-diode oscillator (this is achieved by combining the output of several diodes in the final stage). One avenue of approach for improvement is to use a hybrid tunnel diode-Gunn diode amplifier, in which the tunnel diode input stages significantly reduce the noise figure. Noting that the foregoing applies to gallium arsenide diodes, another avenue of approach is to use indium phosphide devices. The early results with InP Gunn diodes are most encouraging, with noise figures as low as 12 dB reported for amplifiers in the 50- to 60-GHz range.

For reasons identical to those applying to YIG-tuned Gunn oscillators, Gunn amplifiers, be they GaAs or InP, are capable of broad-band operation, 2:1 bandwidth ranges being not unusual. They are greatly superior to IMPATT amplifiers in this respect.

Gunn diode applications Having taken the microwave world more or less by storm, Gunn diode oscillators are widely used and also intensely researched and developed. They are employed frequently as low- and medium-power oscillators in microwave receivers and instruments. The majority of parametric amplifiers now use Gunn diodes as pump sources. They have the advantage over IMPATT diodes of having much lower noise, this being an important criterion in the selection of a pump oscillator. Where very high pump frequencies are required, the technique of using a lower-frequency Gunn oscillator and doubling the frequency with a varactor multiplier is often used.

The higher-power Gunn oscillators (250 to 2000 mW) are used as power output oscillators, generally frequency-modulated, in a wide variety of low-power transmitter applications. These currently include police radar, CW Doppler radar (see Chapter 16), burglar alarms and aircraft rate-of-climb indicators.

12-7 AVALANCHE EFFECTS AND DIODES

In 1958, Read at Bell Telephone Laboratories proposed that the delay between voltage and current in an avalanche, together with transit time through the material, could make a microwave diode exhibit negative resistance. Because of fabrication difficulties and the large amounts of heat that would have to be dissipated, such a diode was not produced until 1965, by Johnston and associates at the same laboratories. The diode was subsequently called the *IMP*act *A*valanche and *T*ransit *T*ime (IMPATT) diode. Two years later, at RCA Laboratories this time, a method of operating the IMPATT

diode that seemed anomalous at the time was discovered by Prager and others. This device, now called the *TRA*pped *P*lasma *A*valanche *T*riggered *T*ransit (TRAPATT) diode, also exhibits negative resistance and holds out a promise of high pulsed powers at the lower microwave frequencies.

12-7.1 IMPATT Diodes

Introduction It was shown in Section 12-5.1 that the tunnel diode has a *dynamic dc negative resistance*. This meant that, over a certain range, current decreased with an increase in voltage, and vice versa. No device has a *static* negative resistance, i.e., with voltage applied one way, and current flowing the other way. This particular point was pursued no further, it being taken for granted that any device which exhibits a dynamic negative resistance *for direct current* will also exhibit it *for alternating current*. If an alternating voltage is applied, current will rise when voltage falls, at an ac rate. We may now redefine negative resistance as *that property of a device which causes the current through it to be* 180° *out of phase with the voltage across it*. The point is important here, because this is the only kind of negative resistance exhibited by the IMPATT diode. One hastens to add that such a negative resistance is quite sufficient. It would not have mattered if the tunnel diode had only this kind of negative resistance (without exhibiting it for dc voltage or current variations)—after all, the oscillations are ac. To summarize; if it can be shown that the voltage current in the IMPATT diode are 180° out of phase, negative resistance in this device will have been proved.

IMPATT diode A combination of delay involved in generating avalanche current multiplication, together with delay due to transit time through a drift space, provides the necessary 180° phase difference between applied voltage and the resulting current in an IMPATT diode. The cross section of the active region of this device is shown in Figure 12-28. Note that it *is* a diode, the junction being between the p^+ and the n layers.

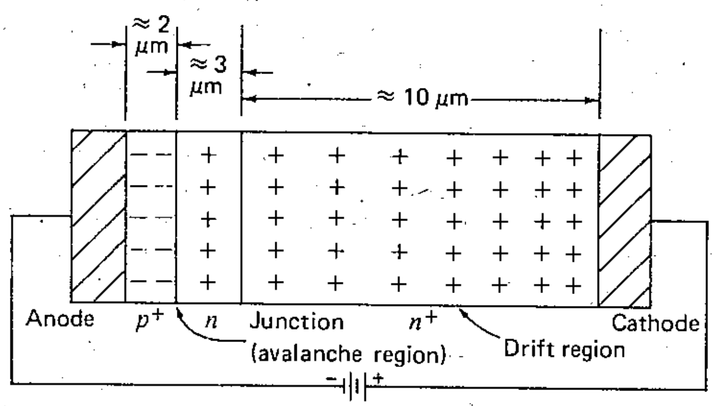

FIGURE 12-28. **IMPATT diode (single-drift) schematic diagram.**

An extremely high voltage gradient is applied to the IMPATT diode, of the order of 400 kV/cm, eventually resulting in a very high current. A normal diode would very quickly break down under these conditions, but the IMPATT diode is constructed so as to be able to withstand such conditions repeatedly. For example, a normal diode breaks down under avalanche conditions because of the enormous powers generated. Consider that the thickness of an IMPATT diode's active region is a few micrometers, to ensure the correct transit time for microwave operation. Its cross-sectional area is similarly tiny, to ensure a small capacitance. With the high-voltage gradient and resulting high current, the power being generated is of the order of *100 MW/cm³*. The delay between the proposal for the IMPATT diode and its first realization was due in no small measure to the problems involved in dissipating such vast amounts of heat. This had to be done, to ensure a satisfactorily low operating temperature for the IMPATT diode, so that it would not be destroyed by melting. Typical operating temperatures of commercial diodes are of the order of 250°C. Such a high potential gradient, back-biasing the diode, causes a flow of minority carriers across the junction. If it is now assumed that oscillations exist, we may consider the effect of a positive swing of the RF voltage superimposed on top of the high dc voltage. Electron and hole velocity has now become so high that these carriers form additional holes and electrons by knocking them out of the crystal structure, by so-called *impact ionization*. These additional carriers continue the process at the junction, and it now snowballs into an avalanche. If the original dc field was just at the threshold of allowing this situation to develop, this voltage will be exceeded during the whole of the positive RF cycle, and avalanche current multiplication will be taking place during this entire time. *Since it is a multiplication process, avalanche is not instantaneous.* As shown in Figure 12-28, the process takes a time such that the current pulse maximum, at the junction, occurs at the instant when the RF voltage across the diode is zero and going negative. A 90° phase difference between voltage and current has been obtained.

The current pulse in the IMPATT diode is situated at the junction. However, it does not stay there. Because of the reverse bias, the current pulse flows to the cathode, at a drift velocity dependent on the presence of the high dc field. The time taken by the pulse to reach the cathode depends on this velocity and of course on the thickness of the highly doped (n^+) layer. The thickness of the drift region is cunningly selected so that the time taken for the current pulse to arrive at the cathode corresponds to a further 90° phase difference. As shown in Figure 12-29, when the current pulse actually arrives at the cathode terminal, the RF voltage there is at its negative peak. Voltage and current in the IMPATT diode are 180° out of phase, and a dynamic RF negative resistance has been proved to exist. Such a negative resistance lends itself to use in oscillators or amplifiers. Because of the short times involved, these can be microwave. Note that the device thickness determines the transit time, to which the IMPATT diode is very sensitive. Unlike the Gunn diode, the IMPATT diode is essentially a narrowband device (especially when used in an amplifier).

Practical considerations Commercial IMPATT diodes have been available for quite some time. They are made of either silicon, gallium arsenide or even indium phosphide. The diodes are mostly mesa, and epitaxial growth is used for at least part of the chip; some have Schottky barrier junctions. Gallium arsenide is theoretically preferable

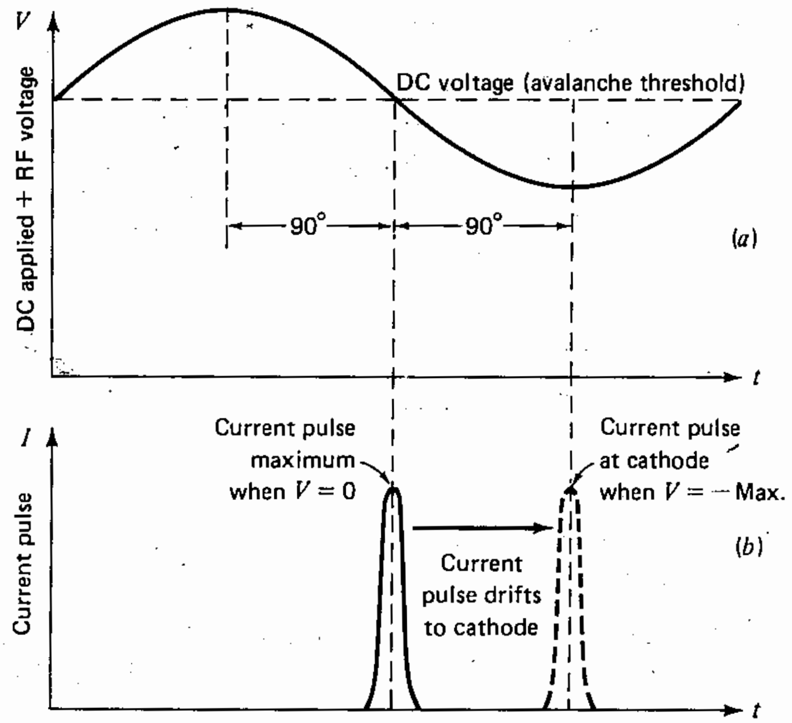

FIGURE 12-29 IMPATT diode behavior. (*a*) Applied and RF voltage; (*b*) resulting current pulse and its drift across diode. (*Note:* Relative size of RF voltage exaggerated.)

and should give lower noise, higher efficiencies and higher maximum operating frequencies. However, silicon is cheaper and easier to fabricate. Accordingly, silicon IMPATT diodes, which came first, are even now preferred for many applications; indeed, it is silicon diodes that currently provide the highest output powers at the highest operating frequencies (in excess of 200 GHz).

The IMPATT diode shown in Figure 12-30 is a typical commercial diode for use below about 50 GHz and could house either a GaAs or an Si chip. At higher frequencies, beam-lead packages almost identical in appearance to those shown in Figure 12-5 tend to be preferred. The construction is deceptively simple. However, a lot of thought and development has gone into its manufacture, particularly the contacts, which must have extremely low ohmic and thermal resistance. Additionally, in a practical circuit, the IMPATT diode is generally embedded in the wall of a cavity, which then acts as an external heat sink.

Until a few years ago, practical IMPATT diodes were unlike Read's original proposal. This called for a double-drift region, whereas Figures 12-28 and 12-30 show diodes with single- (n^+) drift regions. The reason for the initial departure from what was theoretically a higher-efficiency structure was difficulty in fabrication, but this problem has now been solved. For some years IMPATT diodes with two drift regions (one n^+ and the other p^+) have been made commercially. In the manufacturing pro-

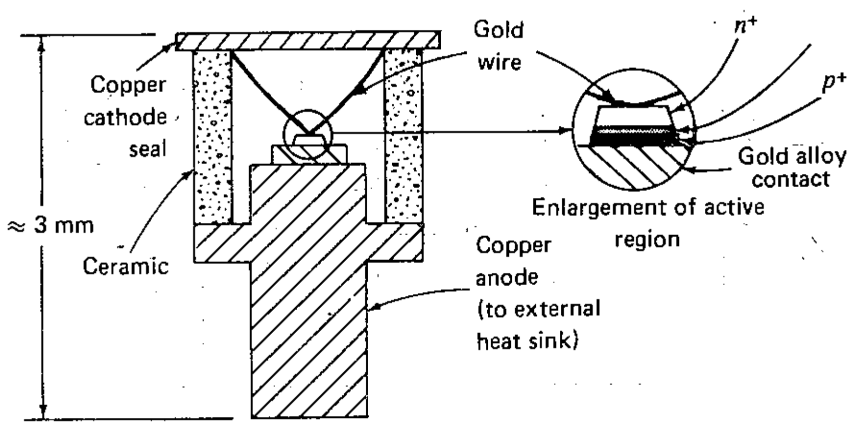

FIGURE 12-30 Typical IMPATT diode.

cess, an n layer is epitaxially grown on an n^+ substrate. The p layer is then grown epitaxially or by ion implantation, and finally the p^+ layer is formed by diffusion. These p^+ -p-n-n^+ devices were at first known as RIMPATT (Read-IMPATT) diodes, but they are now commonly known as double-drift IMPATT diodes. They are undoubtedly the versions used at the highest frequencies and for the highest output powers.

12-7.2 TRAPATT Diodes

The TRAPATT diode is derived from and closely related to the IMPATT diode. Indeed, as pointed out near the beginning of this section, at first it was merely a different, "anomalous," method of operating the IMPATT diode. A greatly simplified operation will now be described.

Basic operation Consider an IMPATT diode mounted in a coaxial cavity, so arranged that there is a short circuit a half-wavelength away from the diode at the IMPATT operating frequency. When oscillations begin, most of the power will be reflected across the diode, and thus the RF field across it will be many times the normal value for IMPATT operation. This will rapidly cause the total voltage across the diode to rise well above the breakdown threshold value used in IMPATT operation. As avalanche now takes place, a plasma of electrons and holes is generated, placing a large potential across the junction, which opposes the applied dc voltage. The total voltage is thereby reduced, and the current pulse is trapped behind it. When this pulse travels across the n^+ drift region of the semiconductor chip, the voltage across it is thus much lower than in IMPATT operation. This has two effects. The first is a much slower drift velocity, and consequently longer transit time, so that for a given thickness the operating frequency is several times lower than for corresponding IMPATT operation. The second point of great interest is that, when the current pulse does arrive at the cathode, the diode voltage is much lower than in an IMPATT diode. Thus dissipation is also much lower, and efficiency much higher. The operation is similar to class C, and indeed the TRAPATT diode lends itself to pulsed instead of CW operation.

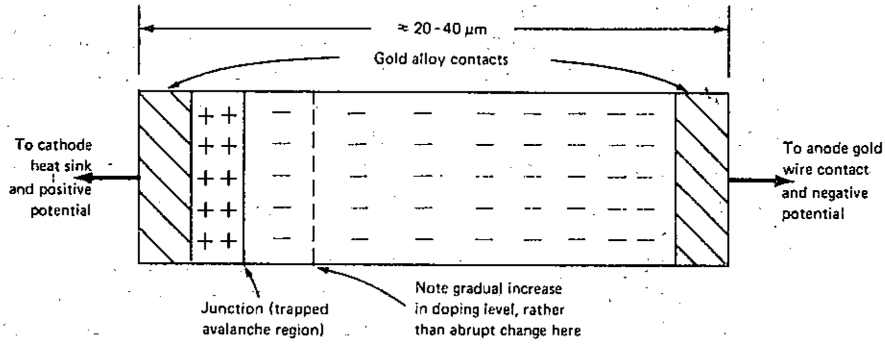

FIGURE 12-31 TRAPATT diode schematic.

Practical considerations Although they were first proposed in the early 1970s, commercial TRAPATT diodes are only now becoming commercially available. They tend to be planar silicon diodes, with structures corresponding to those of IMPATT diodes but with gradual, rather than abrupt, changes in doping level between the junction and the anode. Furthermore, they are likely to use complementary n^+ -p-p^+ structures as shown in Figure 12-31, instead of the p^+ -n-n^+ IMPATT chip of Figure 12-28, for reasons of better dissipation. The two figures should be examined in conjunction with each other.

Because the drift velocity in a TRAPATT diode is much less than in an IMPATT diode, either operating frequencies must be lower or the active regions must be made thinner. In fact, both these considerations are borne out by results obtained. On the one hand, most good experimental TRAPATT results have been for frequencies under 10 GHz, and on the other hand, it has been found that by the time 5 GHz is reached, the width of the depletion layer is only 2 μm. Since the TRAPATT pulse is rich in harmonics, amplifiers or oscillators can be designed to tune to these harmonics, and operation above X band in this manner is possible.

12-7.3 Performance and Applications of Avalanche Diodes

IMPATT diode performance Commercial diodes are currently produced over the frequency range from 4 to about 200 GHz, over which range the maximum output power per diode varies from nearly 20 W to about 50 mW. This means that, above about 20 GHz, the IMPATT diode produces a higher CW power output per unit than any other semiconductor device. Typical efficiency is about 10–20 percent up to 40 GHz, reducing to 1 percent as frequency is raised to 200 GHz. Several diodes' outputs may be combined, giving a significantly greater output. Pulsed powers are generally one magnitude higher. Note that the above figures, for the most part, are for single-drift diodes.

Laboratory devices have produced as much as 30 W CW at 12 GHz, 300 mW at 140 GHz and 75 mW at 220 GHz, with one laboratory reporting 1 mW CW at over 300 GHz. Pulsed powers similarly range from about 50 W at 10 GHz to 3 W at

140 GHz. However, experimental results should be taken with a grain of salt. What is often reported is the best result obtained from several specially made diodes. What is often not reported is that maximum efficiency need not coincide with maximum output power or *that a diode died of thermal runaway soon after the experiment*. It should be noted that results being currently obtained from double-drift IMPATT diodes augur well for the device, especially as regards efficiency, for which figures in excess of 20 percent are being consistently reported, together with higher powers at the highest frequencies.

The biggest problem of IMPATT operation is noise. Avalanche is a very noisy process, and the high operating current helps the generation of shot noise. Thus IMPATT diode oscillators are not as good as either klystrons or Gunn diodes for spurious AM or FM noise, by quite a significant margin. When used as amplifiers, IMPATT diodes produce noise figures of the order of 30 dB, not as good as TWT amplifiers.

IMPATT oscillators and amplifiers The dynamic impedance of an IMPATT diode is $-10\ \Omega$ in parallel with 1 pF, as a good approximation. Like the Gunn diode, therefore, it has a negative resistance which must be placed in a low-impedance environment. Figure 12-32 shows a suitable arrangement. The IMPATT diode is located at the end of the center conductor in a low-impedance coaxial resonator, and a quarter-wave transformer is used to step up the impedance seen at its point of connection. Oscillations are basically at the frequency at which the length of the coaxial resonator is a half-wave, but this is influenced by the capacitance of the varactor diode. This diode is used for tuning, with its capacitance varied by a change in the applied bias. Frequency modulation could be achieved in exactly the same manner. Typical frequency variation is a few hundred megahertz at 10 GHz. Because of their close dependence on transit time through the entire drift space, IMPATT diodes do not lend themselves to tuning over nearly as wide a frequency range as Gunn diodes. Consequently YIG tuning is not used, since varactors match IMPATTs in that regard.

IMPATT diode amplifiers are available with outputs similar to those of oscillators at about the same frequency range. They are comparable to Gunn diode amplifiers in that they also require circulators, but efficiencies for Gunn amplifiers (up to 10 percent) and power outputs are much higher. Gain is similarly 6 to 10 dB per stage, and bandwidths are up to about 10 percent of the center frequency. Higher frequencies of operation, to over 100 GHz, are another attraction, but noise is still a problem.

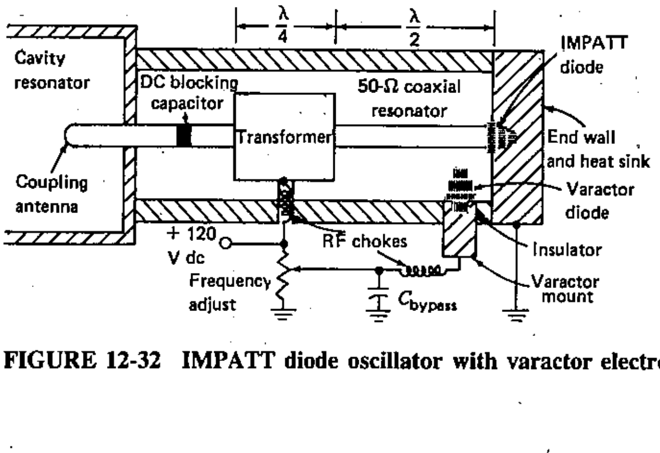

FIGURE 12-32 IMPATT diode oscillator with varactor electronic tuning.

Performance of TRAPATT oscillators and amplifiers As was explained in a preceding section, TRAPATT operation requires a large RF voltage swing, the kind unlikely to be obtained from switching transients. It seems that TRAPATT oscillators most probably start in the IMPATT mode, then switch over when oscillations have built up sufficiently. The circuit must thus be arranged to permit this to happen. However, no such difficulties are encountered with TRAPATT amplifiers, where an adequately large signal is present, being the input. Another practical point which must be taken into account is the extreme TRAPATT sensitivity to harmonics. Thus, when operating in the fundamental mode, care must be taken to ensure that the second, third and even fourth harmonics cannot be maintained in the tuned circuit.

Most TRAPATT oscillators and amplifiers are still in the laboratory stage. However, such impressive results have been obtained that it should not be long before units are available commercially, now that the initial difficulties in establishing and maintaining coherent oscillations seem to have been overcome. With typical duty cycles of the order of 0.1 percent, pulsed powers as high as 200 W at 3 GHz have been produced, with efficiencies in excess of 30 percent. At about 1 GHz, pulsed output powers of 600 W and (independently) an efficiency of 75 percent have been obtained. These figures are for oscillators, but amplifier figures should be comparable. Performance above X band is not very impressive, because of the mode of operation. It should be mentioned that many microwave operations take place at X band or below.

Applications of avalanche diodes IMPATT diodes are more efficient and more powerful than Gunn diodes. However, they have not replaced Gunn diodes, and the reason is mainly their noise and the higher supply voltages needed. It also happens that the majority of low-power microwave oscillator applications can be adequately covered by Gunn diodes, except at the highest frequencies, where they are no match for IMPATTS. However, with the current development in IMPATT and TRAPATT diodes proceeding apace, their use in practical systems is wide and increasing, but they are taking over from low- and medium-power tubes, rather than Gunn diodes. For example, most parametric amplifier designers do not want IMPATTS, because of noise. However, long-distance communications carriers are replacing many of their TWT transmitters with IMPATT ones in microwave links in the large field covered by powers under 10 W. IMPATTs can also eventually replace BWOs and low-power CW magnetrons in several types of CW radar and electronic countermeasures. Finally, when commercial TRAPATT oscillators and amplifiers can produce several hundred watts pulsed, with efficiencies in excess of 30 percent and duty cycles close to 1 percent, a very wide pulsed radar field will be open to them. The first applications here are likely to be in airborne and marine radars.

12-8 OTHER MICROWAVE DIODES

Having discussed in detail the microwave "active" diodes, we are now left with some "passive" microwave diodes to consider. They are passive only to the extent that they are not used in power generation or amplification; apart from that, they are very active indeed in mixers, detectors and power control. The devices in question are the *PIN*, *Schottky-barrier* and *backward* diodes.

12-8.1 PIN Diodes

The PIN diode consists of a narrow layer of p-type semiconductor separated from an equally narrow layer of n-type material by a somewhat thicker region of *intrinsic* material. The intrinsic layer is a lightly doped n-type semiconductor. The name of the diode is derived from the construction (p-intrinsic-n). Although gallium arsenide is used in the construction of PIN diodes, silicon tends to be the main material. The reasons for this are easier fabrication, higher powers handled and higher resistivity of intrinsic region. The PIN diode is used for microwave power switching, limiting and modulation. It was first proposed by R. N. Hall in 1952, and its potential as a microwave switch was first recognized by Uhlir in 1958.

Construction The construction of the PIN diode is shown in Figure 12-33. The advantage of the planar construction is the lower series resistance while conducting. Encapsulation for such a chip takes any of the forms already shown for other microwave diodes. The in-line construction has a number of advantages, including reduced diode shunt capacitance. Also, as shown in Figure 12-33c and d, it lends itself ideally to beam-lead encapsulation, thus interworking excellently with stripline circuits. This construction is often preferred in practice, except perhaps for the highest powers. When fairly large dissipations are involved, the planar construction is better adapted to mounting on a heat sink.

Operation The PIN diode acts as a more or less ordinary diode at frequencies up to about 100 MHz. However, above this frequency it ceases to be a rectifier, because of the carrier storage in, and the transit time across, the intrinsic region. At microwave

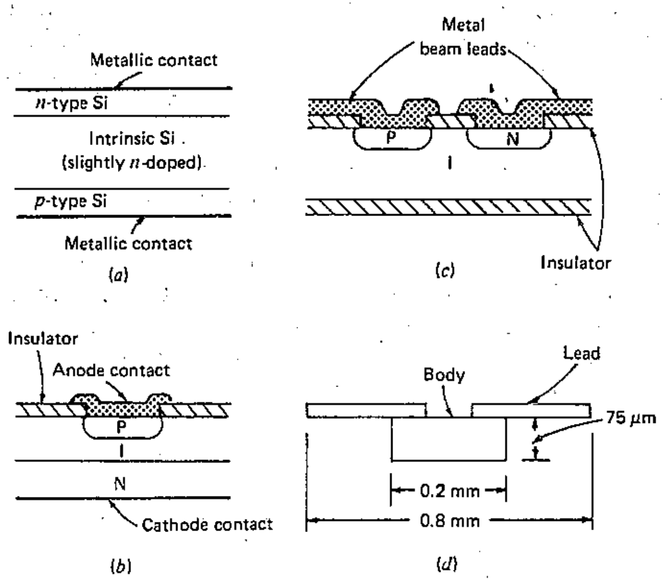

FIGURE 12-33 PIN diode. (a) Schematic diagram; (b) planar diode; (c) planar diode with in-line orientation; (d) beam-lead mounting of in-line diode.

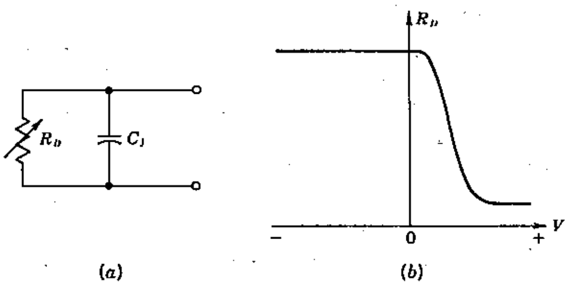

FIGURE 12-34 PIN diode high-frequency behavior. (*a*) Equivalent circuit; (*b*) resistance variation with bias.

frequencies the diode acts as a variable resistance, with a simplified equivalent circuit as in Figure 12-34*a* and a resistance-voltage characteristic as in Figure 12-34*b*.

When the bias is varied on a PIN diode, its microwave resistance changes from a typical value of 5 to 10 kΩ under negative bias to the vicinity of 1 to 10 Ω when the bias is positive. Thus, if the diode is mounted across a 50-Ω coaxial line, it will not significantly load the line when it is back-biased, so that power flow will be unaffected. When the diode is forward-biased, however, its resistance becomes very low, so that most of the power is reflected and hardly any is transmitted. The diode is acting as a switch. In a similar fashion, it may be used as a (pulse) modulator. Several diodes may be used in series or in parallel in a waveguide or coaxial line, to increase the power handled or to reduce the transmitted power in the *OFF* condition.

Performance and applications The applications of the PIN diode are as outlined; they have already been covered in Section 10-5.4, and Figure 10-45 showed a coaxial PIN diode switch. Diodes are available with resistive cutoff frequencies up to about 700 GHz. As for varactor diodes, the operating frequencies do not exceed one-tenth of the above figure. At least one instance of operation at 150 GHz, with specially constructed diodes, has been reported. Individual diodes may handle up to about 200 kW peak (or 200 W average), although typical levels are one magnitude lower. Several diodes may be combined to handle as much as 1 MW peak. Actual switching times vary from approximately 40 ns for high-power limiters to as little as 1 ns at lower powers.

12-8.2 Schottky-Barrier Diode

Schottky junctions have been shown and described throughout this chapter, in conjunction with various devices that use them in their construction (for instance, see Figure 12-4 and its description). Accordingly it will be realized that the Schottky-barrier diode is an extension of the oldest semiconductor device of them all—the point-contact diode. Here the metal-semiconductor interface is a surface—the Schottky barrier—rather than a point contact. It shares the advantage of the point-contact diode in that there are no minority carriers in the reverse-bias condition; that is, there is no significant current from the metal to the semiconductor with back bias. Thus the delay present

in junction diodes, due to hole-electron recombination time, is absent here. However, because of a larger contact area (barrier) between the metal and semiconductor than in the point contact diode, the forward resistance is lower, and so is noise.

The most commonly used semiconductors are "the old faithfuls," silicon and gallium arsenide. As usual, GaAs has the lower noise and higher operating frequency limits; silicon is easier to fabricate and is consequently used at X band and below, in preference to GaAs, *N*-type epitaxial materials are used, and the metal is often a thin layer of titanium surrounded by gold for protection and low ohmic resistance. The device sometimes bears the name *ESBAR* (acronym for epitaxial Schottky-barrier) diode and may also be called the *hot-electron diode*. The latter name is given because electrons flowing from the semiconductor to the metal have a higher energy level than electrons in the metal itself, just as the metal would if it were at a higher temperature. The diodes are encapsulated in any one of the ways already shown for other diodes, with packages corresponding to Figures 12-9 and 12-33*d* common.

Schottky-barrier diodes are available for microwave frequencies up to at least 100 GHz. Like point-contact diodes, they are used as detectors and mixers, mounted as shown previously in Figures 10-44 to 10-46 and discussed in Section 10-5.3. The noise figures of mixers using Schottky-barrier diodes are excellent, rising for as low as 4 dB at 2 GHz to 15 dB near 100 GHz. At frequencies much above X band, GaAs diodes are preferred, since they have lower noise. At the highest frequencies, point-contact diodes are preferred, since they have lower shunt capacitances. For a comparison of Schottky-barrier diode performance with that of other low-noise front ends, see Table 12-2.

12-8.3 Backward Diodes

It is possible to remove the negative-resistance peak and valley region from the tunnel diode of Section 12-5.1, by suitable doping and etching during manufacture. When this is done, the voltage-current characteristic of Figure 12-35 results. This shows the

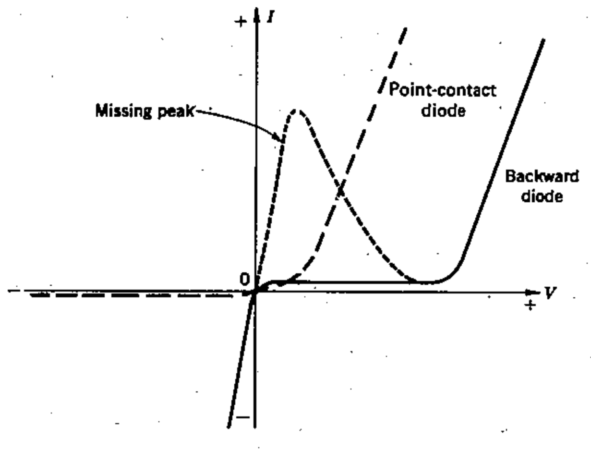

FIGURE 12-35 Backward diode voltage-current characteristic.

rather unusual situation in which, for small applied voltages, the forward current is actually much smaller than the reverse current. The reverse current is large, it will be recalled, because of the very high doping. On the other hand, forward current is low at first because tunneling has been stopped. This diode can therefore be used as a small-signal rectifier. It has the advantage not only of a narrow junction, and therefore a high operating speed and frequency, but also of a current ratio (reverse to forward!) which is much higher than in conventional rectifiers.

When GaAs is used, a maximum signal of about 0.9 V may be applied to the diode before it begins to conduct heavily in the forward direction. This value, although higher than for germanium (silicon is an unsuitable material), is nevertheless quite low. This naturally means that the backward diode is limited, just like the tunnel diode, to lower operating levels. Despite this, the backward diode, or *tunnel rectifier* as it is sometimes called, is in quite common use. Aside from having a high current ratio in the two directions, the backward diode is a low-noise device. It is used in such applications as video detection and low-level mixing, as in Doppler radar (see Section 16-3). Another of its attractions is that it requires a local oscillator signal up to 10 dB lower than that needed by a point-contact diode.

12-9 STIMULATED-EMISSION (QUANTUM-MECHANICAL) AND ASSOCIATED DEVICES

The first *really* low-noise microwave amplifier produced *Microwave Amplification by Stimulated Emission of Radiation;* hence the acronym *maser*. This brand new principle was developed to fruition by Townes and his colleagues in 1954 and provided extremely low-noise amplification of microwave signals by a *quantum-mechanical* process. The *laser,* or optical maser (*l* stands for light), is a development of this idea, which permits the generation or amplification of *coherent* light. In this instance, coherent means single-frequency, in-phase, polarized and directional—just like microwave radio waves. This was also put forward by Professor Townes, in 1958. The overall work was of sufficient importance to make him the 1964 corecipient of the Nobel Prize for physics. The first practical laser was demonstrated by Maiman in 1960.

12-9.1 Fundamentals of Masers

As was found with ferrites in Section 10-5.2, certain materials have atomic systems that can be made to resonate magnetically at frequencies dependent on the atomic structure of the material and the strength of the applied magnetic field. When such a resonance is stimulated by the application of a signal at that frequency, absorption will take place, as in the *resonant absorption* ferrite isolator. Alternatively, emission will occur, if the material is suitably excited, or pumped, from another source. It is upon this behavior that the maser is based.

The material itself may be gaseous, such as ammonia, or solid-state, such as ruby. Ammonia was the original material used, and it is still used for some applications, notably in the so-called *atomic clock* frequency standards. *Extreme* is the correct word to use in describing the stability of such an oscillator. The atomic clock built at

Harvard University in 1960 has a cumulative error which would cause it to be incorrect by only 1 second after more than 30,000 years! From the point of view of microwave amplification, ammonia gas suffered from the disadvantage of yielding amplifiers that worked at only one frequency and whose bandwidth was very narrow. This description will therefore be aimed mainly at the ruby maser.

Fundamentals of operation The electrons belonging to the atoms of a substance can exist in various energy levels, corresponding to different orbit shells for the individual atoms. At a very low temperature, most of the electrons exist in the lowest energy level, but they may be raised by the addition of *specific amounts of energy*. *Quantum theory* shows that a quantum, or bundle of energy, may provide the required energy to raise the level of an electron, provided that

$$E = hf \tag{12-12}$$

where E = energy difference, joules

f = photon frequency, Hz

h = Planck's constant = 6.626×10^{-34} joule · s

Having been excited by the absorption of a quantum, the atom may remain in the excited state, but this is most unlikely to last for more than perhaps a microsecond. It is far more likely that the photon of energy will be reemitted, at the same frequency at which it was received, and the atom will thus return to its original, or *ground,* state. The foregoing assumes, incidentally, that the reemission of energy has been *stimulated* at the expense of absorption. This may be done by such measures as the provision of a structure resonant at the desired frequency and the removal of absorbing atoms, as was done in the original gas maser.

It is also possible to supply energy to these atoms in such quantities and at such a frequency that they are raised to an energy level which is much higher than the ground state, rather than immediately above it. This being the case, it is then possible to make the atoms emit energy at a frequency corresponding to the difference between the top level and a level intermediate between the top level and the ground state. This is achieved by the combination of the previously mentioned techniques (the cavity now resonates at this new frequency) and the application of an input signal at the desired frequency. Pumping thus occurs at the frequency corresponding to the energy difference between the ground and the top energy levels. Reemission of energy is stimulated at the desired frequency, and the signal at this frequency is thus amplified. *Practically no noise is added to the amplified signal*. This is because there is no resistance involved and no electron stream to produce shot noise. The material that is being stimulated has been cooled to a temperature only a few degrees above absolute zero. It now only remains to find a substance capable of being stimulated into radiating at the frequency which it is required to amplify, and low-noise amplification will be obtained.

The original substance was the gas ammonia, while hydrogen and cesium featured prominently among the materials used subsequently. The gaseous substance had the advantage of allowing absorbing atoms to be removed easily. Since the operating frequency was determined very rigidly by the energy levels in ammonia, the range of

frequencies over which the system operated, i.e., its bandwidth, was extremely narrow (of the order of 3 kHz at a frequency of approximately 24 GHz). There was no method whatsoever of tuning the maser, so that signals at other frequencies just could not be amplified. To overcome these difficulties, the traveling-wave ruby maser was invented. This explanation was greatly simplified, especially that of the solid-state maser. Also, some slight liberties with the truth had to be taken in order to present an overall picture that is essentially correct and *understandable*.

The ruby maser A gaseous material is inconvenient in a maser amplifier, as can be appreciated. The search for more suitable materials revealed ruby, which is a crystalline form of silica (Al_2O_3) with a slight natural doping of chromium. Ruby has the advantages of being solid, having suitably arranged energy levels, and being *paramagnetic,* which virtually means "slightly magnetic." This last property is due to the presence of chromium atoms, which have *unpaired electron spins* (like the ferrites of Section 10-5.2). These are capable of being aligned with a dc magnetic field, and this permits not only reradiation of energy from atoms in the desired direction but also some tuning facilities.

Figure 12-36 shows the energy-level situation in a three-level maser, introduced in the previous section. Energy at the correct pump frequency is added to the atoms in the crystal lattice of ruby, raising them to the uppermost of the levels shown (there are many other levels, but they are of no interest here). Normally, the number of electrons in the third energy level is smaller than the number in the ground level. However, as pumping is continued, the number of electrons in level 3 increases until it is about equal to the number in the first level. At this point the crystal saturates, and so-called *population inversion* has been accomplished.

Since conditions have been made suitable for reradiation (rather than absorption) of this excess energy, electrons in the third level may give off energy at the original pump frequency and thus return to the ground level. On the other hand, they may give off smaller energy quanta at the frequency corresponding to the difference between the third and second levels and thus return to the intermediate level. A large number of them take the latter course, which is stimulated by the presence of the cavity surrounding the ruby, which is resonant at this frequency. This course is further aided by the presence of the input signal at this frequency. Since the amount of energy radiated or emitted by the excited ruby atoms at the signal frequency exceeds the energy applied at the input (it does not, of course, exceed the pumping energy), amplification results.

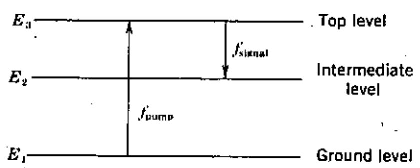

FIGURE 12-36 Energy levels in ruby relevant to maser operation.

The presence of the strong magnetic field (typically about 4 kA/m) has the effect of providing a difference between the three desired energy levels that corresponds to the required output frequency. Any adjustment of this magnetic field will alter the energy levels of the ferrous chromium atoms and therefore provide a form of tuning. This is similar to the situation in ferrites, where it was found that a change in the dc magnetic field changed the frequency of *paramagnetic resonance*. This field strength can be altered to permit the ruby maser to be operated over a frequency range from below 1 to above 6 GHz. For frequencies as high as 10 GHz and above, other materials are often used. *Rutile* is a very common alternative; this is titanium oxide (TiO_2) with a light doping by iron. At the higher frequencies, the required magnetic fields tend to be rather strong, so that the magnet is very often cooled also, to take advantage of *superconductivity* and therefore to give a reduction in the power required to maintain the magnetic field.

In order to consider the effect of cooling the ruby with liquid helium (which is almost invariably done) it is helpful to consider Figure 12-37. Figure 12-37*a* shows the situation at room temperature. Cooling with liquid nitrogen down to only 77 K can also be used, but it results in an increase in noise and a reduction in gain. It is seen that because of the relatively high energy possessed by the electrons at this temperature, quite a number of electrons normally exist in the fourth level, apart from the three so far mentioned. This has the undesirable effect of reducing the number of electrons in the ground level. There are fewer electrons whose energy level can be raised from the first to the third, and consequently fewer electrons that can reradiate their excess energy at the correct frequency. The high temperature is said to *mask* the maser effect. If cooling is applied, the overall energy possessed by the electrons is reduced, as is the number of electrons at the fourth level. As seen in Figure 12-37*b* there are now an adequate number of electrons that can be jumped from the ground to the third level and then down again to the intermediate level. Maser action is maintained. Note that no maser has operated satisfactorily at room temperature. Even if such operation were possible, the noise level would be raised sufficiently to make the noise figure of the maser a very poor second to that of the parametric amplifier.

The noise figure of the cooled ruby maser is governed by the same factors as that of the ammonia maser and is therefore equally low. There is the slight noise due to

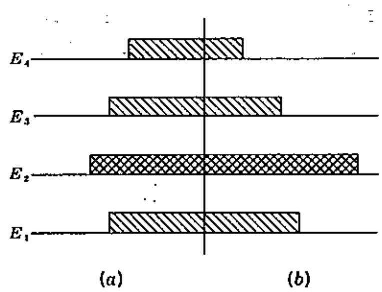

FIGURE 12-37 Energy level populations in suitably pumped ruby. (*a*) At room temperature; (*b*) at liquid helium temperature. (Note the reduction in the fourth-level population in the latter case and the accompanying significant population inversion in levels 2 and 1.)

the random motion of electrons in the ruby (caused by the fact that the temperature of the crystal is above absolute zero). However, most of the noise is due to the associated components, such as the waveguide leading from the antenna, and the noise created at the input to the following amplifier. The first of these problems may be reduced by making the waveguide run as short as possible. This involves mounting the maser at the prime focus of the antenna. Such a solution is practicable only if a Cassegrain or folded horn antenna is used (see Chapter 9), and in fact that is done in practice. The problem of noise from succeeding stages is alleviated in a number of ways. One involves cooling the circulator (which must sometimes be used), in the same way as in a parametric amplifier. It is also possible to increase the gain of the maser, thereby reducing noise reflected from succeeding stages, by making it a two-stage amplifier. The amplifier following the maser can be made a relatively low-noise one, by the use of tunnel diodes or FETs.

12-9.2 Practical Masers and Their Applications

Practical solid-state masers The term *solid-state* is used deliberately here; it does not mean "semiconductor." In terms of the somewhat older maser parlance, it means the opposite of gaseous, i.e., ruby.

The cross section of a ruby cavity maser is shown in Figure 12-38. It is seen to be a single-port amplifier, so that a circulator is needed, just as in so many other microwave amplifiers. In the parametric amplifier, a tuned circuit must be provided for the pump signal as well as for the signal to be amplified. This is not difficult to achieve, but it should be realized that the cavity must be able to oscillate at both frequencies.

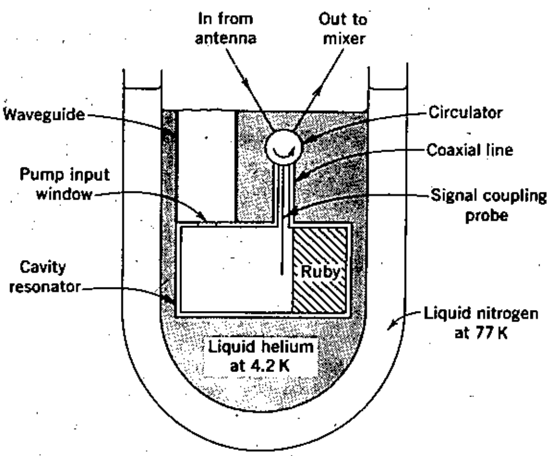

FIGURE 12-38 Schematic diagram of cryogenically cooled ruby maser cavity amplifier (magnet not shown).

From a communications point of view, a disadvantage of the cavity maser is that its bandwidth is very narrow, being governed to a large extent by the cavity itself. It may be typically 1.5 MHz at 1.5 GHz, but some compromise at the expense of gain is possible, noting that the gain-bandwidth product is about 35 MHz. Increasing the bandwidth to even 25 MHz is not practicable, however, since gain by then would not be much in excess of unity.

The solution to the problem is one that has already been encountered a number of times in this chapter; the use of a traveling-wave structure. The resulting operating system is then virtually identical to the one used in the TW paramp. The signal to be amplified now travels along the ruby via a slow-wave structure and grows at the expense of the pump signal. The traveling-wave maser has not only an increased bandwidth but also effectively four terminals, so that a circulator is no longer needed. Such TW masers are used in some older satellite earth stations, built before the subsequent paramp developments.

Performance and applications A typical TW maser operating at 1.6 GHz may have a 25-dB gain, a bandwidth of 25 MHz and a 48-GHz pump requiring 140 mW of CW power. The last two figures are also applicable to the cavity maser, and both types are capable of a noise temperature better than 20 K, i.e., a noise figure better than 0.3 dB. A glance at Table 12-2 will serve as a reminder that the noise performance of masers is unsurpassed.

A disadvantage of the maser is that it is a very low-level amplifier and may saturate for input levels well over 1 μW. While this makes it suitable for radioastronomy and other forms of extraterrestrial communications, radar is a typical application in which a maser could not be used. Not only can much larger radar signals be received in the course of duty, but so can jamming. This would certainly overload a maser RF amplifier, though fortunately without permanent damage. The maser would take about 1 s to recover, during which it would be unusable. Care must be taken not to point the antenna at the ground when a maser amplifier is used, or the ground temperature will create sufficient noise to overload the maser once again.

The parametric amplifier has undergone many improvements in the last decade; therefore the maser is not used as frequently as it once was. Compared to the paramp it is bulkier and more fragile, though somewhat less affected by pump noise or frequency fluctuations. It is narrower in bandwidth and easier to overload, which also means that its dynamic range is not as large. The parametric amplifier has approached the maser's noise performance. The main application for the maser now is in radiotelescopes and receivers used for communications with space probes. Its applications lie where the lowest possible noise is of the utmost importance.

12-9.3 Fundamental of Lasers

As already indicated, the laser is a source of coherent electromagnetic waves at infrared and light frequencies. It operates on principles similar to those of the maser, and indeed an understanding of the maser is virtually a prerequisite to the understanding of its more spectacular stablemate. However, the frequencies are *much* higher; for visible light, these range from 430 to 750 terahertz (THz) (i.e., 430,000 to 750,000 GHz!). It can thus be seen that the scope and information-carrying capacity of lasers is immense.

Introduction The first laser using ruby was proposed in 1958 and demonstrated in 1960, while the first continuously operating laser followed in 1961 and used a mixture of helium and neon gases. Since then, a very large number of other materials have been found suitable, including the other inert gases, argon and krypton, as well as the gallium arsenide and other semiconductor diodes.

Ruby laser The ruby laser is similar to the ruby cavity maser, to some extent, in that stimulation is applied to raise the chromium atoms to a higher energy level to secure a population inversion once again. However, this time pumping is with light, rather than with microwave, energy. Also, no magnetic field is required to modify the existing energy levels because these are already suitable for laser action. The cavity is also different, as can be seen from Figure 12-39. This shows that two parallel mirrors are used, one fully silvered and the other partly so, to enable the coherent light radiation to be emitted through that end. The mirrors must be parallel to a high degree of accuracy and must be separated by a distance that is an exact number of half-wavelengths apart (in the ruby, at the desired frequency). Such an arrangement is called a *Fabry-Perot resonator*. The spiral flash tube pumps light energy into the ruby in pulses, which are generated by the charge and discharge of a capacitor. Cooling is used to keep the ruby at a constant temperature, since quite a lot of the energy pumped into it is dissipated into heat, instead of being radiated as coherent light. Although this cooling also helps laser action, as it did with the maser, room temperature operation is normal.

Pumping raises the electrons to a high energy level, different from that which operated in the maser, since the photon energy is now much higher, because of the higher frequency [this is in accord with Equation (12-2)]. Electrons so raised in energy may fall back either to the ground state, emitting uncoordinated radiation, or else to an intermediate level, as a large number of them do. The energy they lose in the process appears in the form of heat and/or fluorescence. The intermediate level is quasi-stable; electrons remain at it for a few milliseconds, which corresponds to the pumping period. Then their energy rapidly falls to the ground level, with ensuing radiation at the desired frequency. The energy discharge from some of the chromium atoms triggers and coordinates the discharge from the others, with a resulting *correct phase relationship* of all the photons radiated. A large number of these may not escape through the cylindrical sidewalls of the ruby. However, the photons traveling longitudinally are reflected from the silvered end walls and travel back and forth, triggering off other atoms. In this

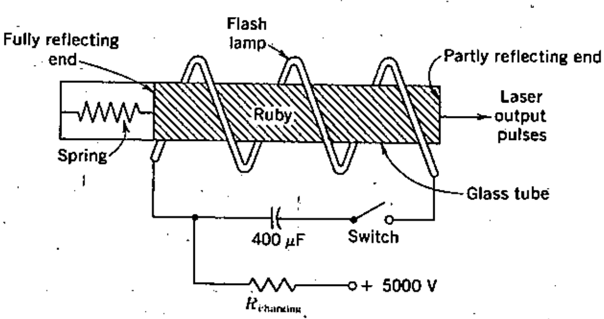

FIGURE 12-39 Basic ruby pulsed laser.

fashion energy builds up, until it is sufficient to escape through the partly silvered end wall, in the form of a very intense short pulse of coherent light that is almost completely *monochromatic* (i.e., single-frequency). The ruby crystal is now in its original state, ready for the next pumping pulse from the flash tube.

The beam of light leaving the ruby crystal is very narrow and almost parallel, with a divergence of less than 0.1°. The frequency spread, or line width, is also very small, of the order of about 1 GHz at a center frequency that is roughly 500,000 GHz (or 500 THz). However, the efficiency is poor (in the vicinity of 1 percent), so that pulsed operation is preferable, in order to permit the dissipated heat to be removed before the next pulse. Cooling also helps, and liquid nitrogen is sometimes used for this. If the chromium doping of the ruby is increased, CW operation becomes possible. The output level is then only milliwatts instead of the megawatts of peak power available with pulsed operation.

It is possible to shorten the pulse duration, without altering the *average* power output of the ruby laser, by the process of *Q-spoiling*, whose effect is to intensify the peak radiated pulse power. In this process, also known as *Q-switching*, one of the ends of the ruby rod is made transparent, and the other is left partly silvered. A mirror is situated behind the unsilvered end, with a shutter placed in front of it. The shutter is closed during pumping, thus preventing laser action and "spoiling" the Q of the Fabry-Perot resonator. This has the effect of greatly helping the population inversion and permits an even larger number of electrons to be situated at the intermediate level. The shutter is opened at the end of the pumping period. With the second mirror now in place, oscillations build up extremely quickly and produce a most intense flash of very short duration; peak powers in excess of 1000 MW are possible.

Two other points should now be raised in connection with solid-state lasers. The first is simply that the laser is an oscillator, unlike the maser. The second is that solid-state lasers are not restricted to using ruby, and other materials have been used to produce other wavelengths. These substances include neodymium, glass doped with gadolinium and the plastic polymethyl methacrylate doped with europium. The last requires ultraviolet pumping and produces a deep crimson light.

12-9.4 CW Lasers and Their Communications Applications

Noncommunications applications of lasers often make the news. They include distance- and speed-measuring equipment, industrial welding, etching (fine enough for the manufacture of integrated circuits) and occasional illumination of a ludicrously small area of the moon with an Earth-operated laser. Everyone has heard of the three-dimensional holograms possible with lasers, while other lasers have been used in optical and other surgery, and still others suggested for military uses. Each of these is a valid laser application, but none of them falls into the category of *communications* applications.

We shall concentrate in this section on those applications of lasers which involve conveying information at a distance. Although it is not essential to have a continuous-wave laser for such work, it does help, and so CW lasers will be the only ones now discussed. Before they are, together with a mention of modulation and detection, it is worth suggesting where they are likely to be used. In fact, it is unlikely that laser

links will ever be used in the same way as microwave links or satellite links. As has often been pointed out, too many things interfere with light in the atmosphere: fog, dust, rain and clouds can all interfere, and so can flying pigeons. It seems that the most spectacular application of laser communications will be in space, while the most frequent workaday one is to send information along optical fibers. This application will be treated in Section 15-2.2—it will be seen that lasers are used along the coaxial cable principles, rather than radio link ones.

Gas lasers The first CW laser, in 1961, was a gas laser using a mixture of helium and neon gases. These are still used, and a simplified He-Ne laser is shown in Figure 12-40. It operates in a manner similar to that of the ruby laser, with the following differences.

1. The mirrors must be as close as possible to being ideally parallel; hence the bellows of Figure 12-40 which are used for fine adjustment.
2. The mirrors must be optically flat, to better than a wavelength, if proper laser action is to take place. This is not as exacting as might at first appear—amateur reflector telescope mirrors are normally ground to an accuracy of one-eighth of a wavelength or better.
3. RF pumping is now required, at a frequency of about 28 MHz for helium-neon. Energy is discharged into the gas mixture via the ring contacts shown.
4. Emission is not at one frequency but at several so-called *lines*. This behavior is due in part to the atomic structure of the gases.
5. Each of the emission lines is extremely pure, having a line width of only a few hertz, each emitted frequency is extremely close to being monochromatic. In practical lasers, gas mixtures provide the narrowest lines, those of solid-state lasers are one magnitude wider and the lines of semiconductor lasers are one magnitude wider still.
6. The beam divergence from parallel is similarly less than in a ruby laser.
7. Such multifrequency oscillation is possible because the dimensions of the resonator (i.e., the distance between the mirrors) are very much greater than a wavelength. The behavior is exactly the same as in a simple oversized cavity resonator, capable of supporting a large number of modes.

Because pumping is continuous, unlike in the solid-state laser, continuous operation is possible. The early gas lasers operated in the infrared region and produced a

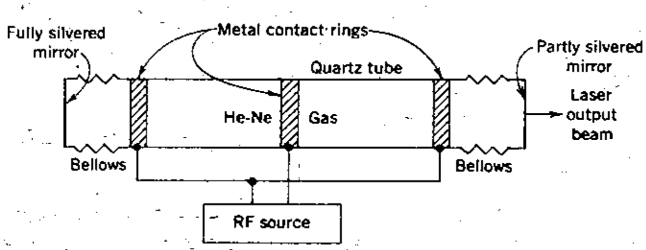

FIGURE 12-40 Schematic diagram of simple CW gas laser. (Note bellows for mirror adjustment; this is the equivalent of cavity tuning.)

few milliwatts with low efficiency. Subsequent improvements have included the use of much shorter tubes to give single rather than multiple lines, laser action with greater efficiency and in the visible spectrum and, more recently, the use of a mixture of carbon dioxide, nitrogen and helium gases. This last device operates in the far infrared spectrum at a wavelength of 10.6 μm, corresponding to a frequency of 28,300 GHz. The process has an efficiency of the order of 20 percent or more, and CW powers as high as 1000 W are possible.

Semiconductor lasers It was discovered in 1962 that a gallium arsenide diode, such as the one shown in Figure 12-41, is capable of producing laser action. This occurs when the diode is forward-biased, so that effective dc pumping is needed (a very convenient state of affairs). Depending on its precise chemical composition, the GaAs laser is capable of producing an output within the range of 0.75 to 0.9 μm, i.e., in the near infrared region (light occupies the 0.39 to 0.77 μm range).

Briefly, the device is an *injection laser,* in which electrons and holes originating in the GaAlAs layers cross the *heterojunctions* (between dissimilar semiconductor materials, GaAlAs and GaAs in this case) and give off their excess recombination energy in the form of light. The heterojunctions are opaque, and the active region is constrained by them to the *p*-layer of GaAs, which is a few micrometers thick, as shown. The two ends of the slice are very highly polished, so that reinforcing reflection takes place between them as in other lasers, and a continuous beam is emitted in the direction shown. The laser is capable of powers in excess of 1 W, which is far higher

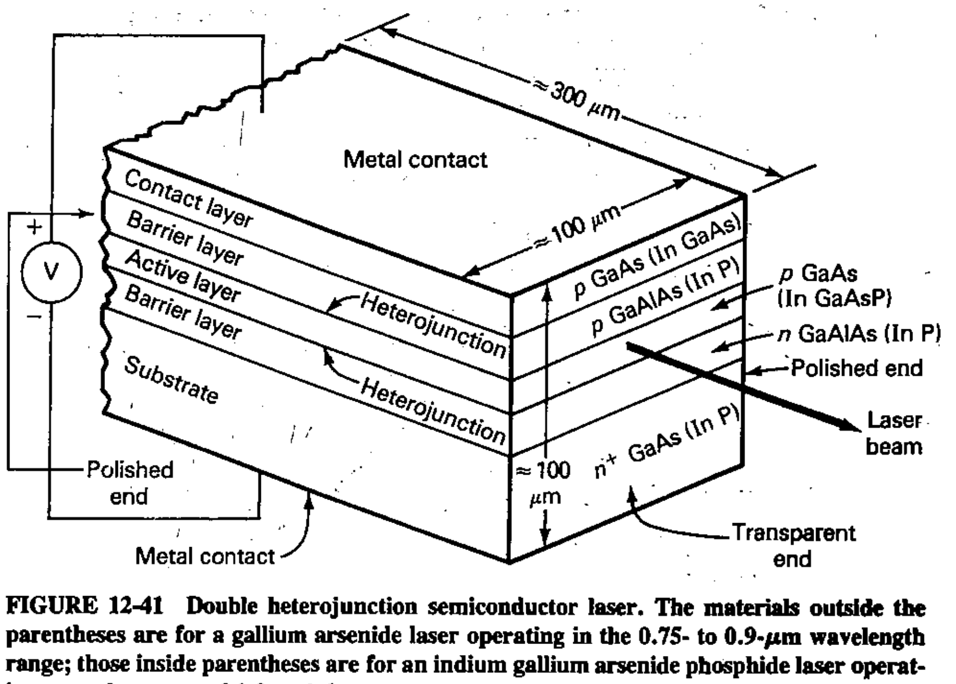

FIGURE 12-41 Double heterojunction semiconductor laser. The materials outside the parentheses are for a gallium arsenide laser operating in the 0.75- to 0.9-μm wavelength range; those inside parentheses are for an indium gallium arsenide phosphide laser operating over the range of 1.2 to 1.6 μm.

than the 1 mW, or so, necessary to send along optic fibers, as will be seen in Chapter 18.

The indium gallium arsenide phosphide laser, also illustrated in Figure 12-41, is a much more recent development than the GaAs device, having been evolved during the late 1970s. The motive force was a desire to produce laser outputs at wavelengths longer than those which the GaAs laser is capable of producing, to take advantage of "windows" in the transmission spectrum of optic fibers—these are discussed in more detail in Chapter 18. Consequently, the InGaAsP lasers are less well developed at the time of writing, and so many of the world's optic fiber communications systems still operate at wavelengths of about 0.85 μm, whereas, as will be shown in Chapter 18, transmissions at wavelengths of 1.3 or 1.55 μm incur significantly less attenuation than at 0.85 μm in optic fibers. By the early to mid-1980s, the teething problems with the new laser materials were being solved, and all new lightwave systems were being designed for wavelengths of 1.3 μm or greater.

12-9.5 Other Optoelectronic Devices

Although *light-emitting diodes* and *photodiodes* are not quantum-mechanical devices, they are semiconductor devices closely associated with lasers. It is most convenient to cover them here.

Light-emitting diodes (LEDs) The construction of an LED is similar to that of a laser diode, as indeed is the operational mechanism. Once again electrons and holes are injected across heterojunctions, and light energy is given off during recombination. The materials used are the same as for the corresponding laser diodes, but the structure is simpler, there are no polished ends and laser action does not take place. Consequently, power output is lower (perhaps one-twentieth) than for the laser, a much wider beam of light results and the light itself is no longer monochromatic. A small lens is often used to couple the output of the LED to the optic fiber.

Despite the foregoing, the LED does have a number of advantages over the laser. For example, it is a good deal cheaper and tends to be more reliable. Moreover, the LED, unlike the laser, is not temperature-sensitive, so that it can operate over a large temperature range without the need for elaborate temperature control circuits which the laser may require. In practice, lasers tend to be used in a fairly large proportion of practical systems, especially the more exacting ones, noting that pulse modulation is normally used, and the light output of lasers can be pulsed at much higher rates than that of LEDs.

Photodiodes A PIN diode, such as any of the ones shown in Figure 12-33, is capable of acting as a photodiode. If a large reverse bias, of the order of 20 V or more, is applied to such a diode, no current will flow. However, if the diode absorbs light quanta through a window on the *p* side, each quantum will cause an electron-hole pair to be created in the intrinsic depletion layer, and a corresponding current will flow in the external circuit. Within limits, this current will be proportional to the intensity of the impinging light, so that photodetection is taking place.

The original photodiode semiconductor was germanium, and it is still used for wavelengths in excess of about 1.1 μm; for shorter wavelengths silicon is preferred. Because of the well-known sensitivity of germanium to temperature, research is currently taking place among the newer semiconductor materials, such as GaAlAs and InGaAs, to find a replacement for the germanium PIN photodetector.

Avalanche photodiodes (APDs) A problem with the PIN photodiode is that it is not overly sensitive; no gain takes place in the device, in that a single photon cannot create more than one hole-electron pair. This problem is overcome by the use of the avalanche photodiode, which, in some respects, operates in a manner similar to the IMPATT diode.

An APD, such as the one shown in Figure 12-42, is operated with a reverse voltage close to break-down. Like the IMPATT, the APD is capable of withstanding sustained break-down. As in the PIN photodetector, a light quantum impinging on the diode will cause a hole-electron pair to be created, but this time avalanche multiplication can take place, as in the IMPATT, so that the initial electron-hole pair will cause several others to be created, with consequently increased current flowing through the eternal circuit. The extent of avalanche multiplication can be gauged from the fact that a typical APD is 10 to 150 times more sensitive than a PIN photodetector.

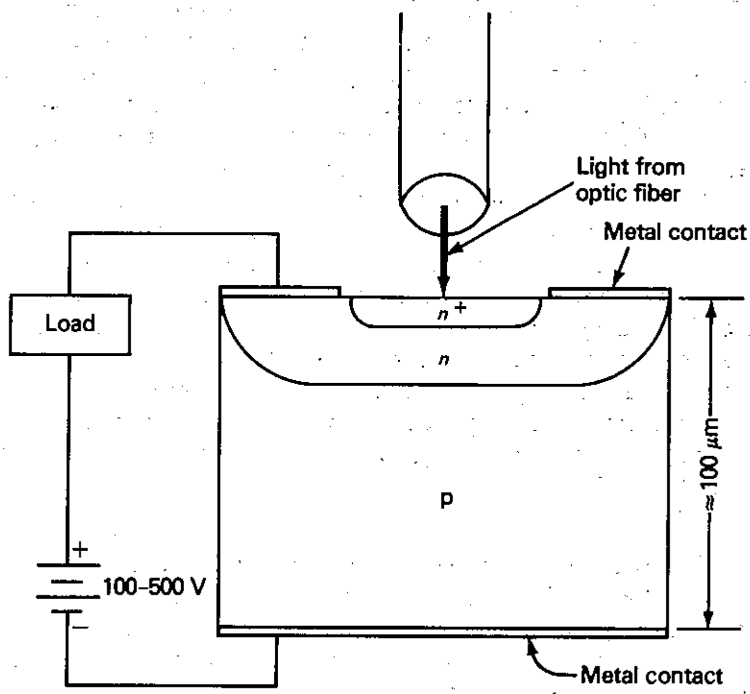

FIGURE 12-42 Avalanche photodiode construction and schematic. (Note similarity to IMPATT diode schematic in Fig. 12-28.)

The materials used for APDs are the same as for the corresponding PIN diodes. Because the voltage gradient across the APD is so high, electron and hole drift is higher than for the PIN diode, and the response time is similarly faster, typically 2 nS compared with 5 nS for the PIN diode. It follows that the APD can be used for higher pulse modulation rates than the PIN. There is a fairly close correlation between light transmitters and receivers in fiber-optic systems. Those less exacting systems which use LEDs for transmission are also likely to use PIN photodiodes for reception. The systems requiring higher sensitivities and higher modulation bit rates are likely to use lasers for transmission and avalanche photodiodes for reception.

MULTIPLE-CHOICE QUESTIONS

Each of the following multiple-choice questions consists of an incomplete statement followed by four choices (a, b, c, and d). Circle the letter preceding the line that correctly completes each sentence.

1. A parametric amplifier must be cooled
- **a.** because parametric amplification generates a lot of heat
- **b.** to increase bandwidth
- **c.** because it cannot operate at room temperature
- **d.** to improve the noise performance

2. A ruby maser amplifier must be cooled
- **a.** because maser amplification generates a lot of heat
- **b.** to increase bandwidth
- **c.** because it cannot operate at room temperature
- **d.** to improve the noise performance

3. A disadvantage of microstrip compared with stripline is that microstrip
- **a.** does not readily lend itself to printed circuit techniques
- **b.** is more likely to radiate
- **c.** is bulkier
- **d.** is more expensive and complex to manufacture

4. The transmission system using two ground planes is
- **a.** microstrip
- **b.** elliptical waveguide
- **c.** parallel-wire line
- **d.** stripline

5. Indicate the *false* statement. An advantage of stripline over waveguides is its
- **a.** smaller bulk
- **b.** greater bandwidth
- **c.** higher power-handling capability
- **d.** greater compatibility with solid-state devices

6. Indicate the *false* statement. An advantage of stripline over microstrip is its
- **a.** easier integration with semiconductor devices
- **b.** lower tendency to radiate
- **c.** higher isolation between adjacent circuits
- **d.** higher Q

7. Surface acoustic waves propagate in
- **a.** gallium arsenide
- **b.** indium phosphide
- **c.** stripline
- **d.** quartz crystal

8. SAW devices may be used as
- **a.** transmission media like stripline
- **b.** filters
- **c.** UHF amplifiers
- **d.** oscillators at millimeter frequencies

9. Indicate the *false* statement. FETs are preferred to bipolar transistors at the highest frequencies because they

a. are less noisy
b. lend themselves more easily to integration
c. are capable of higher efficiencies
d. can provide higher gains

10. For best low-level noise performance in the X-band, an amplifier should use
a. a bipolar transistor
b. a Gunn diode
c. a step-recovery diode
d. an IMPATT diode

11. The biggest advantage of the TRAPATT diode over the IMPATT diode is its
a. lower noise
b. higher efficiency
c. ability to operate at higher frequencies
d. lesser sensitivity to harmonics

12. Indicate which of the following diodes will produce the highest pulsed power output:
a. Varactor
b. Gunn
c. Schottky barrier
d. RIMPATT

13. Indicate which of the following diodes does not use negative resistance in its operation:
a. Backward
b. Gunn
c. IMPATT
d. Tunnel

14. One of the following is *not* used as a microwave mixer or detector:
a. Crystal diode
b. Schottky-barrier diode
c. Backward diode
d. PIN diode

15. One of the following microwave diodes is suitable for very low-power oscillators only:
a. Tunnel
b. avalanche
c. Gunn
d. IMPATT

16. The transferred-electron bulk effect occurs in
a. germanium
b. gallium arsenide
c. silicon
d. metal semiconductor junctions

17. The gain-bandwidth frequency of a microwave transistor, f_T, is the frequency at which the
a. alpha of the transistor falls by 3 dB
b. beta of the transistor falls by 3 dB
c. power gain of the transistor falls to unity
d. beta of the transistor falls to unity

18. For a microwave transistor to operate at the highest frequencies, the (indicate the *false* answer)
a. collector voltage must be large
b. collector current must be high
c. base should be thin
d. emitter area must be large

19. A varactor diode may be useful at microwave frequencies (indicate the *false* answer)
a. for electronic tuning
b. for frequency multiplication
c. as an oscillator
d. as a parametric amplifier

20. If high-order frequency multiplication is required from a diode multiplier,
a. the resistive cutoff frequency must be high
b. a small value of base resistance is required
c. a step-recovery diode must be used
d. a large range of capacitance variation is needed

21. A parametric amplifier has an input and output frequency of 2.25 GHz, and is pumped at 4.5 GHz. It is a
a. traveling-wave amplifier
b. degenerate amplifier
c. lower-sideband up-converter
d. upper-sideband up-converter

22. A nondegenerate parametric amplifier has an input frequency f_i and a pump frequency f_p. The idler frequency is
a. f_i
b. $2f_i$
c. $f_i - f_p$
d. $f_p - f_i$

• 23. Traveling-wave parametric amplifiers are used to
a. provide a greater gain
b. reduce the number of varactor diodes required

   c. avoid the need for cooling
   d. provide a greater bandwidth

24. A parametric amplifier sometimes uses a circulator to
   a. prevent noise feedback
   b. allow the antenna to be used simultaneously for transmission and reception
   c. separate the signal and idler frequencies
   d. permit more efficient pumping

25. The nondegenerate one-port parametric amplifier should have a high ratio of pump to signal frequency because this
   a. permits satisfactory high-frequency operation
   b. yields a low noise figure
   c. reduces the pump power required
   d. permits satisfactory low-frequency operation

26. The tunnel diode
   a. has a tiny hole through its center to facilitate tunneling
   b. is a point-contact diode with a very high reverse resistance
   c. uses a high doping level to provide a narrow junction
   d. works by quantum tunneling exhibited by gallium arsenide only

27. A tunnel diode is loosely coupled to its cavity in order to
   a. increase the frequency stability
   b. increase the available negative resistance
   c. facilitate tuning
   d. allow operation at the highest frequencies

28. The negative resistance in a tunnel diode
   a. is maximum at the peak point of the characteristic
   b. is available between the peak and valley points
   c. is maximum at the valley point
   d. may be improved by the use of reverse bias

29. The biggest advantage of gallium antimonide over germanium for tunnel-diode use is that the former has a
   a. lower noise
   b. higher ion mobility
   c. larger voltage swing
   d. simpler fabrication process

30. Negative resistance is obtained with a Gunn diode because of
   a. electron transfer to a less mobile energy level
   b. avalanche breakdown with the high-voltage gradient
   c. tunneling across the junction
   d. electron domains forming at the junction

31. For Gunn diodes, gallium arsenide is preferred to silicon because the former
   a. has a suitable empty energy band, which silicon does not have
   b. has a higher ion mobility
   c. has a lower noise at the highest frequencies
   d. is capable of handling higher power densities

32. The biggest disadvantage of the IMPATT diode is its
   a. lower efficiency than that of the other microwave diodes
   b. high noise
   c. inability to provide pulsed operation
   d. low power-handling ability

33. The magnetic field is used with a ruby maser to
   a. provide sharp focusing for the electron beam
   b. increase the population inversion
   c. allow room-temperature operation
   d. provide frequency adjustment

34. The ruby maser has been preferred to the ammonia maser for microwave amplification, because the former has
   a. a much greater bandwidth
   b. a better frequency stability
   c. a lower noise figure
   d. no need for a circulator

35. Parametric amplifiers and masers are similar to each other in that both (indicate *false* statement)
   a. must have pumping
   b. are extremely low-noise amplifiers
   c. must be cooled down to a few kelvins
   d. generally require circulators, since they are one-port devices

36. A maser RF amplifier is not really suitable for
a. radioastronomy
b. satellite communications
c. radar
d. troposcatter receivers

37. The ruby laser differs from the ruby maser in that the former
a. does not require pumping
b. needs no resonator
c. is an oscillator
d. produces much lower powers

38. The output from a laser is monochromatic; this means that it is
a. infrared **c.** narrow-beam
b. polarized **d.** single-frequency

39. For a given average power, the *peak* output power of a ruby laser may be increased by
a. using cooling
b. using Q spoiling
c. increasing the magnetic field
d. dispensing with the Fabry-Perot resonator

40. Communications lasers are used with optical fibers, rather than in open links, to
a. ensure that the beam does not spread
b. prevent atmospheric interference
c. prevent interference by other lasers
d. ensure that people are not blinded by them

41. Indicate the *false* statement. The advantages of semiconductor lasers over LEDs include
a. monochromatic output
b. higher power output
c. lower cost
d. ability to be pulsed at higher rates

REVIEW PROBLEMS

1. A microwave signal has a purely resistive output impedance of 500 Ω, and its load is matched for maximum power transfer. A negative resistance is now placed across the circuit, turning it into an amplifier. If the value of this negative resistance is $-200\ \Omega$, what will be the power gain of the amplifier?

2. If, in Problem 12-1, the load and source resistance are now both 1000 Ω, what must be the value of the negative resistance to give a power gain of 23 dB?

REVIEW QUESTIONS

1. With the aid of appropriate sketches, describe basic stripline and microstrip circuits. From what previously studied transmission media are they derived?

2. What are the advantages and disadvantages of stripline and microstrip with respect to waveguides and coaxial transmission lines? What are the conditions under which waveguides and coax would be preferred?

3. What are the applications of microstrip and stripline circuits? Which is the more convenient to use in hybrid MICs? Why?

4. Discuss the construction and applications of surface acoustic wave devices, illustrating the answer with a sketch of a typical SAW component.

5. Discuss the high-frequency limitations of transistors, comparing and contrasting them with those of vacuum tubes.

6. Illustrating your answer with sketches, describe the construction of microwave bipolar and field-effect transistors.

7. Compare the performance and general construction of hybrid and monolithic MICs.

8. Discuss the performance and applications of microwave transistors and MICs, illustrating your answer with graphs of power output and noise versus frequency.

9. With the aid of suitable sketches, discuss the materials, construction and characteristics of microwave varactors.

10. Discuss briefly the basic theory of varactor frequency multipliers. Define the term *nonlinear capacitance*.

11. Discuss the capabilities and applications of varactor and snap-recovery diode frequency multipliers.

12. What is a parametric amplifier? Discuss its fundamentals *in full*, and state the ways in which it differs from an orthodox amplifier.

13. Describe the *nondegenerate* negative-resistance parametric amplifier. Show a simple circuit of this device, and explain the function of the *idler* circuit.

14. What is the most common type of very low-noise parametric amplifier? Show the block diagram of such a device, explaining carefully the function of the circulator. Does the use of the circulator have any drawbacks? Can its use be avoided?

15. Draw the circuit diagram of a representative TW parametric amplifier, and briefly explain how it works. Why must the pump frequency be not *too* much higher than the signal frequency in this type of amplifier?

16. Discuss the noise performance of parametric amplifiers and the factors influencing it. Why is *cryogenic* cooling sometimes used? Is it compulsory? What are the advantages of *not* cooling cryogenically?

17. Discuss the advantages and list the applications of parametric amplifiers. Contrast the applications of paramps cooled by various means with those of uncooled ones.

18. Using energy-band (Fermi-level) diagrams, explain the tunnel-diode characteristic (voltage-current curve) point by point. Take it for granted that quantum-mechanical tunneling will take place under favorable conditions.

19. Discuss the problems connected with the biasing of a tunnel diode and their solution. Illustrate the discussion with a practical tunnel-diode circuit.

20. Explain why it is possible to obtain amplification by using a device which exhibits negative resistance.

21. Discuss the performance, advantages and applications of tunnel-diode amplifiers, and then compare them, in turn, with each of the other microwave low-noise amplifiers.

22. What is the significant and very important difference between the *Gunn effect* and all the other properties of semiconductors?

23. Explain fully the Gunn effect, whereby negative resistance, and therefore oscillations, are obtainable under certain conditions from bulk gallium arsenide and similar semiconductors. Why are Gunn devices called *diodes?*

24. Sketch a Gunn diode construction, and describe it briefly. What are some of the performance figures of which Gunn diodes are capable?

25. What are Gunn *domains?* How are they formed? What do they do?

26. How does the domain formation in a Gunn diode respond to the tuning of the cavity to which the diode is connected? Sketch a cavity Gunn oscillator.

27. Describe the construction, fabrication and encapsulation of Gunn diodes.

28. Discuss the performance and operation of (*a*) YIG-tuned Gunn oscillators; (*b*) Gunn diode amplifiers.

29. What do the acronyms *IMPATT* and *TRAPATT* stand for?

30. What are the applications of Gunn oscillators and amplifiers?

31. Draw the schematic diagram of an IMPATT diode, and fully explain the two effects that combine to produce a 180° phase difference between the applied voltage and the resulting current pulse.

32. Show an encapsulated IMPATT diode, and discuss some of the practical considerations involved. What is a double-drift IMPATT diode?

33. Briefly describe the basic operating mechanism of TRAPATT diodes, using a suitable sketch. Why is the drift through this diode much slower than through a comparable IMPATT diode? What implications does this have for the operational frequency range of the TRAPATT diode?

34. Compare the performance of IMPATT and TRAPATT oscillators with that of Gunn oscillators and amplifiers. Consider also their relative applications.

35. What is the major drawback of avalanche devices? What limitations does this place on their applications?

36. With the aid of a suitable sketch, describe the construction of a PIN diode. What does PIN stand for? Briefly explain the operation of this diode.

37. Discuss the performance and applications of Schottky-barrier diodes, and list the competitors for those applications.

38. Write a survey of semiconductor diode and bulk effect microwave generators, describing briefly the construction, operation, performance and applications of each.

39. How does the backward diode differ from the tunnel diode? What is this device used for?

40. What is a maser? What does its name signify? What application does it have?

41. Discuss the fundamentals of the maser, and explain the various levels at which electrons may be found, the connection between the pumping frequency and these levels and finally what is done to make the electrons reemit the energy they receive from the pump source, instead of absorbing it. Why is the maser such a low-noise device?

42. Show the energy levels in a ruby crystal relevant to maser operation. What is meant by the terms *population inversion* and *saturation?* How does the presence of the magnetic field affect the situation? What else can the magnetic field be used for?

43. From what point of view is cooling of a ruby maser with liquid helium preferable to cooling with liquid nitrogen? Discuss the causes of noise in a maser amplifier, and describe some of the steps taken in practice to reduce it.

44. What are the capabilities and performance of the maser?

45. Discuss fully the operation of the ruby laser. Show a basic sketch of one.

46. What are the outstanding characteristics of the ruby laser? Describe the process of *Q-spoiling* and its function. What is the big disadvantage of this laser from a communications point of view?

47. Compare and contrast the operation and applications of the gas laser with those of the ruby laser.

48. Briefly explain the operation of a semiconductor laser, using a sketch showing the construction of this device.

49. What is the major application of semiconductor lasers? How do GaAs and InGaAsP devices compare in this regard?

50. How does the performance of light-emitting diodes compare with that of semiconductor lasers? What are their respective applications?